

Db 209 AVHTA-----TENIPOTPOKIMMIMPGVGVDEWGTGCDNTPLHADYEW---RYTPLE 261
 QY 197 --GGEGGSGPFTLDWT-----DNDFPFDSRRG-RGDWTFDG 230
 Db 262 ALDEESGDNEEPVEVEEPEADEEVSVPISGAIYFTFENFDIWSIAHGW-NG 320
 QY 231 NRYDLT--DKNYSRDGMILLATLRKQSEFNGQVPRDDEPA 270
 Db 321 QMFNATWYDSQITFSNGIMRFAIDK-----DDEEA 351

RESULT 2
 Q45648 PRELIMINARY; PRT; 276 AA.
 AC 045648;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DE LICHENASE PRECURSOR.
 GN BGAL.
 OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 NCBI_TaxID=1409;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=137;
 RX MEDLINE=94288605; PubMed=7517127;
 RA Taberner C., Coll P.M., Fernandez-Abalos J.M., Perez P.,
 RA Santamaria R.I.;
 RT "Cloning and DNA sequencing of bgaA, a gene encoding an endo-beta-1,3-
 RT RT 1,4-glucanase, from an alkalophilic Bacillus strain (N137).";
 RL Appl. Environ. Microbiol. 60:1213-1220(1994).
 DR EMBL: Z12151; CA878135.1;
 DR HSSP: P23904; IAJK.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS: PR00737; GLHYDRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KM Signal.
 FT SIGNAL. 1 31 POTENTIAL.
 FT CHAIN 32 276 LICHENASE.
 SQ SEQUENCE 276 AA; 31770 MW; D047F8A34CA9EBE2 CRC64;

Query Match 13.6%; Score 247.5; DB 2; Length 276;
 Best Local Similarity 36.1%; Pred. No. 1.9e-11;
 Matches 61; Conservative 20; Mismatches 75; Indels 13; Gaps 6;
 QY 30 FSGAELTYLEVOYGKFEARKMAAGTVSMFLYONGSEINDGRPWVEVDIEVGKNP 89
 Db 96 YKGEIRLTDNFYHGLFEVSMKPAKSTGVSSFTY-TGFWDENDWDIEDIEFGKDT 154
 QY 90 GSFQSNITGKGAOKTSEKHNHVAVSPADQAFHTYGLEMTPNVVRMTVDQEVKTEGGQ 149
 Db 155 TKIQFNYFNGVCG--NHHYHELGFADADDFVTYAFEMRPESIRNRYVNGELVHTA---- 207
 QY 150 VSNLTGT-OGLRFNLMSSESA-AMVGQFDESKLPLEOFIMWVYKYP 196
 Db 208 TENIPOTPOKIMMIMPGVGVDEWGTGCDNTPLHADYEW---RYTPLE 253

RESULT 3
 Q45691 PRELIMINARY; PRT; 801 AA.
 AC 045691;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE FAMILY 11 XYLANASE /FAMILY 16 BETA (1,3-1,4) GLUCANASE.
 GN XYNL.
 OS Ruminococcus flavefaciens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Ruminococcus.
 NCBI_TaxID=1265;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=17;
 RA Aurilia V., Martin J.C., Munro C.A., Mercer D.K., Flint H.J.;
 RT "Organisation and strain distribution of genes responsible for the
 RT utilization of xylans by the rumen cellulolytic bacterium Ruminococcus
 RT flavefaciens 17";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ132472; CAB51934.1;
 DR HSSP: P23904; IAJK.
 DR InterPro: IPR000757;
 DR InterPro: IPR001137;
 DR InterPro: IPR003305;
 DR Pfam: PF00457; Glyco_hydro_11; 1.
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR Pfam: PF02018; CBD_6; 1.
 DR PRINTS: PR00911; GLHYDRLASE11.
 DR PRINTS: PR00737; GLHYDRLASE16.
 DR PROSITE: PS00776; GLYCOSYL_HYDROL_F11_1; 1.
 DR PROSITE: PS00777; GLYCOSYL_HYDROL_F11_2; 1.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 FT INIT_MET 0
 FT SEQUENCE 801 AA; 88887 MW; 9FTB837389F8F5C2 CRC64;

Query Match 13.3%; Score 242.5; DB 2; Length 801;
 Best Local Similarity 34.9%; Pred. No. 1.8e-10;
 Matches 58; Conservative 21; Mismatches 74; Indels 13; Gaps 6;
 QY 30 FSGAELTYLEVOYGKFEARKMAAGTVSMFLYONGSEINDGRPWVEVDIEVGKNP 89
 Db 633 YSGEFTNTNFYHGYECGMOAMKNDGVSSFTYTGPS---DNDPWEIDIEILGKNT 689
 QY 90 GSFQSNITGKGAOKTSEKHNHVAVSPADQAFHTYGLEMTPNVVRMTVDQEVKTEGGQ 149
 Db 690 TQGFNYFTNGQKH--ELTYLGFDFSSSEAHYTGQDPNVIAMVYDGRREYRA---- 742
 QY 150 VSNLTGTG-LRFNLMSSESA-AMVGQFDESKLPLEOFIMWVYK 193
 Db 743 TQDIPKTPGKIMMAMPGLTVDMLKAFN-GRTPLTAHYQWYVYK 787

RESULT 4
 Q45691 PRELIMINARY; PRT; 239 AA.
 AC 045691;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE ENDO-BETA-1,3-1,4-GLUCANASE.
 GN BEG1.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 NCBI_TaxID=1423;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 8565;
 RA van Rensburg P., Van Zyl W.H., Pretorius I.S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U60810; AAB05759.1;
 DR HSSP: P27051; IGBG.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS: PR00737; GLHYDRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 SQ SEQUENCE 239 AA; 26969 MW; C4219760D13F878A CRC64;

Query Match 13.1%; Score 238.5; DB 2; Length 239;

Best Local Similarity 33.7%; Pred. No. 7.8e-11;
Matches 59; Conservative 27; Mismatches 72; Indels 17; Gaps 7;

OY 23 TNVSKDPSGAEIYLYEVOYGFEPARKMAAASGVSMFLYONGSEIADGRPVEVDI 82
DB 77 TSPYNNKFDCEGNSVQYVYGLTEVRKPKAKNGIVSFFTYTGPE--GTPDEIDI 133
OY 83 EVLGNKPGSFQSNITITGKAGAKTSEKHAHVSPA--ADQAFHTYGLEWTPNYVAVTWVG 139
DB 134 EFLGKDKTKKQVFNYYTNA-----ENHEKLADLGFDAAANHYHTAFWQPNSTIKWYVDW 187
OY 140 QEVKRTGEGVSNLTGTQGLRFNLMSESA--WVGQFDESKLPLFQFINWVKVYK 193
DB 188 Q-LKHTATQTQIPARGK--IMMNLNMGVSDWLGSYNGVN-PLVAYHDWVRKYK 238

RESULT 5
OYFDC9 PRELIMINARY; PRT; 212 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ENDO-1,3-1,4-BETA-GLUCANASE (FRAGMENT).
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1406;

RN [1]
RP SEQUENCE FROM N.A.
RA Yao W., Wang Y., Song W., Yang K., Su Z.;
RT "Gene cloning of an antifungal protein."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DB EMBL: AF284449; AAC02415.1; -
FT NON_TER 1 1
FT SEQUENCE 212 AA; 24127 MW; 7E707EFBE7A440 CRC64;

Query Match Best Local Similarity 12.7%; Score 232.5; DB 2; Length 212;

Matches 62; Conservative 26; Mismatches 73; Indels 17; Gaps 9;

OY 20 ALTTNVSADPSGAEIYLYEVOYGFEPARKMAAASGVSMFLYONGSEIADGRPVE 79
DB 48 SLTSSAVANK-EDGGEYSKNTYRGIVVNMKPKAKNGIVSFFTYTGP--ANGTQWDE 103
OY 80 VDIEVLGNKPGSFQSNITITGKAGAKTSEKHAHVSPAADQAFHTYGLEWTPNYVAVTWVG 139
DB 104 IDIEFLGDKTKKQVFNYYTNGIGH--EKVVDLGFDAASGFHTYAFWQPNSTIKWYVDG 160
OY 140 QEVKRTGEGVSNLTGTQGLRFNLMSESA--WVGQFDESKLPLFQFINWVKVYK 195
DB 161 -VLKHT--ATTNIPKPTPGQIMMNLNMGVSDWLGSYNGVN-PLVAYHDWV--KYP 210

RESULT 6
OYFDC9 PRELIMINARY; PRT; 237 AA.

AC 007856; 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA-(1,3-1,4)-GLUCANASE PRECURSOR.
OS Streptococcus bovis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1315;
RN [1]
RP SEQUENCE FROM N.A.
RA Ekinci M., Flint H.J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z92911; CAB07443.1; -
DR HSSP: P23904; IAKK.
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLHTRDLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
FW SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT SEQUENCE 237 AA; 26989 MW; 7DEF5BC53790470 CRC64;

Query Match Best Local Similarity 12.6%; Score 230; DB 2; Length 237;
Matches 54; Conservative 26; Mismatches 69; Indels 14; Gaps 6;

OY 30 FSGAEIYLYEVOYGFEPARKMAAASGVSMFLYONGSEIADGRPVEVDIEVLGNP 89
DB 81 YTGGEWRSKREYGLFQVNMKPKAKNGIVSFFTYTGP--DQTKWDEIDIEFLGDT 137
OY 90 GSFQSNITITGKAGAKTSEKHAHVSPAADQAFHTYGLEWTPNYVAVTWVGQEVKRTGEGQ 149
DB 138 TKVQFNYYT---SGQGNHEIYLYNLFSDASQGFHTYGFPMQADHTIYVDGRAVITA----- 190
OY 150 VSNLTGTQGLRFNLW--SSESAAMVQGFDESKLPLFQFINWV 189
DB 191 YNNIPSTPGKIMMNAWPGTHEVDSWLGAVN-GRPELVAYYDWI 232

RESULT 7

ID 014412 PRELIMINARY; PRT; 245 AA.

AC 014412; 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN LICA.

OS Orpiniomyces sp. (strain PC-2).
OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
OC Neocallimastixaceae; Orpiniomyces.
OX NCBI_TaxID=37164;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 30-41.
RX MEDLINE=97464427; PubMed=9324248.

RA Chen H., Li X.-L., Ljungdahl L.G.;

RT "Sequencing of a 1,3-1,4-beta-D-glucanase (lichenase) from the anaerobic fungus Orpiniomyces strain PC-2: properties of the enzyme expressed in Escherichia coli and evidence that the gene has a bacterial origin."

RT J. Bacteriol. 179:6028-6034(1997).

RL J. Bacteriol. 179:6028-6034(1997).

CC 1- FUNCTION: HYDROLYSES B-GLUCANS CONTAINING MIXED BETA-1,3 AND BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; OPTIMAL TEMPERATURE IS

CC 45 DEGREES CELSIUS AND THE OPTIMUM PH IS 5.8-6.2.

CC 1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES

CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.

CC 1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.

DB EMBL: U6813; AAD04192.1; -

DR HSSP: P23904; IAKK.

DR InterPro: IPR000757; -

DR Pfam: PF00722; Glyco_hydro_16; 1.

DR PRINTS: PR00737; GLHTRDLASE16.

DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; FALSE_NEG.

KW Hydroxylase; Glycosidase; Signal.

FT SIGNAL 1 29

FT CHAIN 30 245 BETA-GLUCANASE.

FT ACT_SITE 134 134 NUCLEOPHILE (BY SIMILARITY).

FT ACT_SITE 138 138 PROTON DONOR (BY SIMILARITY).

FT DISULFD 63 90 BY SIMILARITY.

FT CONFLICT 36 36 S -> L (IN AA SEQUENCE).

SO SEQUENCE 245 AA; 27929 MW; 08B5DF57D89F2DCD CRC64;

Query Match 10.4%; Score 190.5; DB 3; Length 245;
 Best Local Similarity 31.6%; Pred. No. 3.0e-07;
 Matches 54; Conservative 17; Mismatches 87; Indels 13; Gaps 6;

OY 21 LTTNVSADSGALTYLEEVQYKFEARKMAASGVSSMFLYQNGSEIADGRPWVEV 80
 DB 79 LTTDRDGGSTGCGYRKNYKMGFOVNMKPIKPNGVSSFFTYTGPS--DGTKDEI 135
 OY 81 DIEVLGNPSPFOSNITIGKAGAKTSEKHNHVASPADQAFHTYGLFMTPNVRYRWTDGO 140
 DB 136 DIEFLGVDYTKVQENYNTNGGHH---EHIIYLGFDASQGHHTYGLFMRARSTIWTYDGT 192
 OY 141 EVKRTGEGQVSNLTGTGCG-LRENLMSSEA-AWVGQFDESKLPLFOFINWV 189
 DB 193 AVYTA----YDNIPDTPGKIMNMAMNGIGVDMLRPFN-GRTNISAYYDNY 238

RESULT 8

O9K7X6 PRELIMINARY; PRT; 302 AA.

ID O9K7X6; AC O9K7X6; DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE HYBRID-ENDO-BETA-1,3-1,4 GLUCANASE.
 GN BH3231.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 NCBI_TaxID=8665;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / JCM 9153;
 RA Takami H., Nakase K., Takai Y.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF001518; BAB06950.1; -
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro.16; 1.
 DR PRINTS: PR00737; GLHYDRASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL.F16; UNKNOWN.1.
 SO SEQUENCE 302 AA; 33975 MW; 2EDDA630FDS355CA CRC64;

Query Match 10.4%; Score 190.5; DB 2; Length 302;
 Best Local Similarity 27.8%; Pred. No. 5e-07;
 Matches 62; Conservative 34; Mismatches 98; Indels 29; Gaps 9;

OY 29 DFGAGELYLEEVQYKFEARKMAASGVSSMFLYQNGSEIADGRPWVEVDIEVLGN 88
 DB 16 NYCEAEYSTNHFYGYGLYEVSMKPSNVSGVSSFFTYTGPS--YNGAPMDEIDIEFLGND 73
 OY 89 PGFOSNITIGKAGAKTSEKHNHVASPADQAFHTYGLFMTPNVRYRWTDGOEVRKTEGG 148
 DB 74 TTVQVQRYNYNGG---NEILYDLGFDANSENNTYAFDMQEHNTYVNGELVATATEN 130
 OY 149 OVSNLTGTGGLRENLMSSE-----SAWVGQFDESKLPLFOFINWVYKYP-----GQG 109
 DB 131 IPSN---PSKMMNINNTYCIDEMAGAYWQNMNAS-----YEW---RYTNNENNSQT 178
 OY 200 EGGSDFTLMDTDFDPSGRW--GKGDWTFDGNRVDLTDKNI 240
 DB 179 PLASDFOLHACEYSDARGVSMQGVGSF-YPGSWIKFDVNL 220

RESULT 9

O9FI31 PRELIMINARY; PRT; 282 AA.

ID O9FI31; AC O9FI31; DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE XYLOGLUCAN ENDO-1,4-BETA-D-GLUCANASE.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;

RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=9937451; PubMed=10470850;
 RA Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
 RT Sequence features of the regions of 1,011,550 bp covered by seventeen
 RT pl and TAC clones."
 RL DNA Res. 6:183-195(1999).
 DR EMBL: AB017064; BAB11071.1; -
 SO SEQUENCE 282 AA; 32437 MW; FBB8C587F6C9A3C3 CRC64;

Query Match 10.1%; Score 185; DB 10; Length 282;
 Best Local Similarity 27.4%; Pred. No. 1.2e-06;
 Matches 59; Conservative 34; Mismatches 98; Indels 24; Gaps 9;

OY 24 NVSAKDFSGALTYLEEVQYKFEARKMAA--ASGVSSMFLYQNGSEIADGRPWVEVD 81
 DB 53 SLSDKFGSGGFSHQEFLYKVEVQMKLVPGNSAGTYFYFK-----SPGTYWDEID 106
 OY 82 IEVLGK---NPGSFOSNITIGKAGAKTSEKHNHVASPADQAFHTYGLFMTPNVRYRWTD 138
 DB 107 FEEFGNIGSHPTLHTNWVY-KGTGDEKQFHLMEPTVY--PHTCYIINPQVAVFTID 163
 OY 139 GOEVRKTEGGQVSNLTGTGGLRF--NLMSSEA-WVG--OFDESKLPLFOFINWV 191
 DB 164 GIPREFKNEALGVPRPKQPMRLVSLWEAEHWATRGLENTDMSKAFPTAFYRYNV 223
 OY 192 YKTPGQEGSGDFTLMDTDFDPSGRW--GKGDWTFDGNRVDLTDKNI 240
 DB 224 DACVWSNGKSSCSANSSWFTQVLDLFGKKNRVKMAQ 258

RESULT 10

O42800 PRELIMINARY; PRT; 302 AA.

ID O42800; AC O42800; DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE RASP F 9 (FRAGMENT).
 GN RASP F 9.
 OS Aspergillus fumigatus (Sartorya fumigata).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 NCBI_TaxID=5085;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 42202;
 RA Hermann S.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 42202;
 RX MEDLINE=98141802; PubMed=9482698;
 RA Cramer R.;
 RT "Recombinant Aspergillus fumigatus allergens: from the nucleotide
 RT sequences to clinical applications."
 RT Int. Arch. Allergy Immunol. 115:99-114(1998).
 DR EMBL: AJ223327; CA11266.1; -
 DR HSSP: P23904; IAJK.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro.16; 1.
 KW Allergen.
 FT NON_TER 1 1
 FT 302 302

```

SQ      SEQUENCE 302 AA: 32317 MW: 86676a20066fc5a4 cgc64:

Query Match          9.8%, Score 178; DB 3; Length 302;
Best Local Similarity 24.5%; Pred. No. 4.5e-06;
Matches 82; Conservative 41; Mismatches 100; Indels 112; Gaps 15;

OY      3 IKKTAVKSAALVAAAAAATTNVSAKDF-----SCAELYTELEVO-----42
DB      38 LEKTTDPNNKGLASALTYTADFTSASALDQEWYAGVPGVPGQCAETFAVKQGDAPITIDF 97
OY      43 ---YGFEPARMKMAASCTVSMPLFYONGSEIADRPWVEVDIEVLGKPKGSFOSNIITG 99
DB      98 YFFPKAEVYVMAAAGCTGVSSIVLESDDL-----EVDMEVLGGDTTOVQINRYF-G 148
OY      100 KAGAAKTSEKHHAVSPADQ-AFHTTYGLEWTPPNVYRWTVDQGEVR---KTEGQVSNL 153
DB      149 K-GDPTFYDRGYTVVPATPQETFFHTYTTIDMTKDAVTWSDGAVYRLLTYNDAKGGTRFPQ 207
OY      154 TGTGGLRRNINWSESAA-----WVGQF-DESKLPLQFIWVYVYKTKTPQGGEGSDP 205
DB      208 TPRM-LRIGSWGAGDPSPNPKGTIEWAGGLTDSAGPYTMVYKSVRIENAPAE-----259
OY      206 TLDMTDFNDFPDGSRWKGQDW--TFDGNRYVLTLTKNITYSRDGMILIALTRGQSEFNQ 262
DB      260 STTYSDN-----SGSMOSIKFDGS-VDI-----281
OY      263 VPRDEPAPQSSSSAPASSSSVYPASSSSVYPASSSS 297
DB      282 -----SSSSSVTSSSTTYPASSASS 300

RESULT 11
OY      080803      PRELIMINARY;      PRT;      282 AA.
AC      080803:
DT      01-NOV-1998 (TREMBLrel. 08. Created)
DT      01-NOV-1998 (TREMBLrel. 08. Last sequence update)
DT      01-MAY-2000 (TREMBLrel. 13. Last annotation update)
DE      T8F5.9 PROTEIN.
GN      T8F5.9.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC      Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. COLUMBIA:
RA      Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Liu S.,
RA      Li J., Arunjo R., Au M., Brendel V., Buehler E., Conway A.B.,
RA      Conway A.R., Dewar K., Feng J., Kim C., Kutz D., Li Y., Palm C.J.,
RA      Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,
RA      Theologis A.;
RT      "Arabidopsis thaliana chromosome 1 BAC T8F5 complete sequence.";
RL      Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. COLUMBIA:
RA      Theologis A.;
RL      Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. COLUMBIA:
RA      Theologis A.;
RL      Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. COLUMBIA:
RA      Theologis A.;
RL      Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AC004512; AAC27142.1; -.
DR      HSSP; P23904; IACK.

```

```

DR Mendel:31461; AArch:1017;31461.
DR InterPro:IPR000757;-
DR Pfam:PF00722; Glyco_hydro_16; 1
SQ SEQUENCE 282 AA; 31992 MW; 00F9732C0C44172D CRC64;

Query Match 9.5%; Score 173; DB 10; Length 282;
Best Local Similarity 25.6%; Pred. No. 1e-05;
Matches 64; Conservative 36; Mismatches 92; Indels 58; Gaps 12;

QY 24 NVSAKDFSGAEIYLTLEEYOGKFEARMMAA-ASGTVSSMELYONGSEIDGRPWVEVD 81
   ::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 53 SLSDKSSGSGFQSGNQDFELYKGAEPYQMKLYLPGNSGYLTFTFLK-----SPGTWDELD 106
QY 82 IEVLCK---NPGSFOSNIITGKAGAKTSEKHAAVSPADAQFHTYGLEMPYNYRWTVD 138
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 107 FEFLLNIGSHHYTLHTLNYVT-KGTGDKKQDFELMDEPVTN-FHYCYCTWNPQRIETFD 163
QY 139 GQEVKTKGGGVQSNL-TGTGGLRF-NLMSESAAMV--QPEKSLPLPQFINWVKV 191
   ::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 164 GIPLEFNPPEIIGVPPPTROPRIYASLMEHEHWATRGLEKTMWSKAPPTAF----- 217
QY 192 KYPTGQGEQEGSDFLFDITDNEFDPEGSRWKGMDTFDGNRYDLTDKNIYSHDGLIAL 251
   ::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 218 YR-----NPNV-DGCVMANGRKSSCANSWFLQK-----L 246
QY 252 TRKGQSFNG 261
Db 247 DSNGTTRMK 256

RESULT 12
Q9WXN1 PRELIMINARY; PRT; 642 AA.
AC Q9WXN1;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE LAMINARINASE.
GN TM0024.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=MSB8 / DSM 3109;
RC MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Hat D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001690; AAD35118.1; -.
DR HSSP; P23904; IAKK.
DR TIGR; TM0024; -.
DR InterPro; IPR003305; -.
DR Pfam; PF02018; CBD_6; 2.
SQ SEQUENCE 642 AA; 72540 MW; 6431C9825597BF7E CRC64;

Query Match 9.5%; Score 173; DB 2; Length 642;
Best Local Similarity 23.4%; Pred. No. 2.9e-05;
Matches 70; Conservative 49; Mismatches 118; Indels 62; Gaps 16;

QY 12 LAVAAAAAALTNNVSAKDFSGAEIYLTLEEYOGKFEARMMAASGVSSMFLYONGS 69
   ||::||::||::||::||::||::||::||::||::||::||::||::||
Db 265 LVIEARKQGVSDDEGTYYTSARMTTEKFEIKYKIDIRAKLRGKGIWPAIMWL--GN 322
QY 70 EIAD-GRWV-EVDI-EVLGNKPGSFOSNI-TGKAGAKTSEKHAAVSPADAQFHT 123

```

```

Db 323 NIEVGPPTCGEIDIMELHDTRTVYGTAGHPGYSGASIGVAYHLPREGVDPFSEDFHI 382
QY 124 YGLEMTPNVYRMTVDGQEVKRTGEGOVSNLTGQGLRFLNLMSSSE-----SAAW 172
Db 383 FSIEMENDEVEWYVDQQLHYVLSKDELAEL-----GLE---WVFDHPPELLINAVGQYMP 435
QY 173 GQFDE-SKLPLEQFINWVVKYK-----YTPGEGEGSDFTLMDTDFNPDFTF 216
Db 436 GYDPEFTTQFQRMVYIDYIRYKMDNPETITGEVDDCEYEQAOQAGREYVEEQINN-GTF 494
QY 217 DGS-----RWKGDWTFDGNRVDLTDKNYRSDGMLLALTRKGOSEFNGQ 262
Db 495 DEPIVNDQANNPDEWFIWQAGDYGISGARV--SDYGV--TDGYAYTITADPGDTWHIO 549

```

```

RESULT 13
Q9MOD1 PRELIMINARY; PRT; 277 AA.
ID 09MOD1
AC 09MOD1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE XYLOGLUCAN ENDO-1, 4-BETA-D-GLUCANASE-LIKE PROTEIN.
GN AT4G30290.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL161576; CAB81022.1;
DR InterPro: IPR000757;
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; UNKNOWN.1.
SQ SEQUENCE 277 AA; 31563 MW; B801F67E0CD72E53 CRC64;

```

```

Query Match 9.5%; Score 172.5; DB 10; Length 277;
Best Local Similarity 26.6%; Pred. No. 1.1e-05;
Matches 54; Conservative 34; Mismatches 74; Indels 41; Gaps 9;

```

```

QY 24 NVSAKDFSGAELTYLLEVOYGFKEARMKMAA--ASGVSSMFLYONGSEIADGRPVVEVD 81
Db 48 SLSLDKSSSGSGROSEFLGKAEVOMKLVPGNSAGTIVTFYLK-----SPETTDEID 101
QY 82 IEVLK--NPGSFQSNITTKAGAKTSEKHNHVASPADQAFHTGYLETPNYVRYMTVD 138
Db 102 FEELGNIISHGPTLHTNVT-KSGDGKEQOFLMPPTAN--FHTYCTIMNPQRIIFTVD 158
QY 139 GQEVKRTGEGOVSNLTGQGLR-----NLMSSESAANVG--QFDESKLPLEFOF 185
Db 159 GJPIREFMNAE-----SRGVPEPTKQPMRLVSLMEAEHMTATRGILETDMSKAFETAY 212
QY 186 IN-----WVKVYKTPGQGE 200
Db 213 YRNYNEGCVWVNGKSCVCPANSQ 235

```

```

RESULT 14
Q9MOD1 PRELIMINARY; PRT; 646 AA.
ID 060039
AC 060039;
DT 01-NOV-1996 (Tremblrel. 01, Created)

```

```

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE LAMINARINASE (EC 3.2.1.39) (GLUCAN ENDO-1,3-BETA-D-GLUCOSIDASE) ((1-
DE >3)-BETA-GLUCAN ENDONHROLASE) (ENDO-1,3-BETA-GLUCANASE).
GN LJCA.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxId=2337;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=42706-MC24;
RA MEDLINE=93356813; PubMed=8352795;
RA Dakhova O., Kurepina N., Zverlov V., Svetlichny I.,
RA Vellikdovorskaya G.;
RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana
RT genes coding for enzymes of carbohydrate substrate degradation.";
RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=42706-MC24;
RX MEDLINE=97312008; PubMed=9168619;
RT Zverlov V.V., Volkov I.Y., Vellikdovorskaya T.V., Schwarz W.H.;
RT Highly thermostable endo-1,3-beta-glucanase (laminarinase) Lama from
RT Thermotoga neapolitana: nucleotide sequence of the gene and
RT characterization of the recombinant gene product.";
RL Microbiology 143:1701-1708(1997).
CC -1 CATALYTIC ACTIVITY: HYDROLYSIS OF 1,3-BETA-D-GLUCOSIDIC LINKAGES
IN 1,3-BETA-D-GLUCANS.
DR EMBL: Z47974; CAB88008.1;
DR HSSP: P14090; IULO.
DR InterPro: IPR000757;
DR InterPro: IPR003305;
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR Pfam: PF02018; Cbf_6_2.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 646 AA; 73332 MW; 7FE335509284D75 CRC64;

```

```

Query Match 9.4%; Score 171; DB 2; Length 646;
Best Local Similarity 23.4%; Pred. No. 4.1e-05;
Matches 70; Conservative 50; Mismatches 117; Indels 62; Gaps 16;

```

```

QY 12 LAVAAAAAALTYNVSAKDFSGAELTY--LEVOYGFKEARMKMAAASGVSSMFLYONGS 69
Db 269 LVTEARKEQVSDRYGYDTSARITTEGKFEIKYKIEIRAKLPKKGIMPALMML--GN 326
QY 70 EIND-GRPV--EVDI-EVLGKNDS-FQSNITTKAGAKTSEKHNHVASPAD--QAFHT 123
Db 327 NIEVGPPTCGEIDIMELHDTRTVYGTAGHPGYSGASIGVAYHLPREGVDPFSEDFHV 386
QY 124 YGLEMTPNVYRMTVDGQEVKRTGEGOVSNLTGQGLRFLNLMSSSE-----SAAW 172
Db 387 FSIEMENDEVEWYVDQQLHYVLSKDELAEL-----GLE---WVFDHPPELLINAVGQYMP 439
QY 173 GQFDE-SKLPLEQFINWVVKYK-----YTPGEGEGSDFTLMDTDFNPDFTF 216
Db 440 GYDPEFTTQFQRMVYIDYIRYKMDNPETITGEVDDCEYEQAOQAGREYVEEQINN-GTF 498
QY 217 DGS-----RWKGDWTFDGNRVDLTDKNYRSDGMLLALTRKGOSEFNGQ 262
Db 499 DEPIVNDQANNPDEWFIWQAGDYGISGARV--SDYGV--TDGYAYTITADPGDTWHIO 553

```

```

RESULT 15
Q9MOD2 PRELIMINARY; PRT; 282 AA.
ID 09MOD2
AC 09MOD2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE XYLOGLUCAN ENDO-1, 4-BETA-D-GLUCANASE-LIKE PROTEIN.
GN AT4G30280.
OS Arabidopsis thaliana (Mouse-ear cress).

```

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL161576; CAB81021.1; -
 DR InterPro; IPR000757;
 DR Pfam; PF00722; Glyco_hydro_16; 1.
 DR PRINTS; PR00737; GLHYDRLASE16.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; UNKNOWN_1.
 SO SEQUENCE 282 AA; 32072 MW; 437B3AC683F71B31 CRC64;

Query Match 9.3%; Score 170.5; DB 10; Length 282;
 Best Local Similarity 27.6%; Pred. No. 1.6e-05;
 Matches 59; Conservative 35; Mismatches 97; Indels 23; Gaps 10;
 OY 24 NVSAKDFSGAEIYLTLEEVYQGFAPARKMAA--ASGTIVSMFLYQNGSEIADGRPWVEVD 81
 Db 53 SLSDKSSSGFQSGNOEFLYKAEVQMKLVPGNSAGVTTFYLK-----SPGTTWDEID 106
 OY 82 IEVIGK--NPGSFQSNITIGKAGAKTSEKHHAVSPADQAFTITGLEMTPNIVRWTV 138
 Db 107 FEFIGNLSGHEYLATHTVYT-KGSGDKEQDFHLMEDPTVN--FHTYCTITWNPORIIFTVD 163
 OY 139 GQEVKRTGGGVSNL--TGTOGLRF--NLMSSESAAWV--OPDESKLPLFOFINWKV 191
 Db 164 GIPIREKNSSEISGVPPPTKQPMRLYASLWEAEHWATRGLEKTDWSKAPPTAFYRNYV 223
 OY 192 YKYPGQEGGSDFTLDM-TDNPTFDGSRWKG 224
 Db 224 EGCYVANGKSSCPANSSWFTQQLDS-NGQTRMKG 256

Search completed: August 23, 2001, 11:16:06
 Job time: 148 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:17:23 ; Search time 45.21 Seconds

(without alignments)
467.989 Million cell updates/sec

Title: US-09-654-652a-3

Perfect score: 349
Sequence: 1 MNIKRTVAKSLALVAAAAA.....AKGAKVNPNGHKRYVNFEEH 349

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 412676 seqs, 60623988 residues

Word size: 0

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: /SIDS8/gcgdata/geneseq/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	2.9	144	21	Arabidopsis thaliana
2	10	2.9	209	21	Arabidopsis thaliana
3	10	2.9	268	21	Arabidopsis thaliana
4	10	2.9	319	21	Arabidopsis thaliana
5	10	2.9	322	21	Arabidopsis thaliana
6	10	2.9	333	21	Arabidopsis thaliana
7	10	2.9	384	21	Arabidopsis thaliana
8	10	2.9	387	21	Arabidopsis thaliana
9	9	2.6	337	22	Hairy protein amin
10	8	2.3	33	18	Arctic fish antifr
11	8	2.3	37	13	Antifreeze protein

12	8	2.3	37	20	AAV23879
13	8	2.3	37	20	AAV23880
14	8	2.3	37	20	AAW86157
15	8	2.3	37	20	AAW86158
16	8	2.3	37	20	AAW86156
17	8	2.3	37	21	AAV44713
18	8	2.3	38	12	AAV14375
19	8	2.3	38	20	AAV25427
20	8	2.3	40	11	AAV08077
21	8	2.3	41	11	AAV08080
22	8	2.3	41	11	AAV08082
23	8	2.3	41	11	AAV08086
24	8	2.3	41	11	AAV08097
25	8	2.3	43	20	AAV24058
26	8	2.3	46	11	AAV08098
27	8	2.3	50	17	AAV08078
28	8	2.3	51	11	AAV08078
29	8	2.3	52	11	AAV08081
30	8	2.3	52	11	AAV08083
31	8	2.3	62	11	AAV08079
32	8	2.3	66	11	AAV08067
33	8	2.3	73	11	AAV08068
34	8	2.3	82	20	AAV23877
35	8	2.3	97	20	AAV86160
36	8	2.3	114	20	AAV60122
37	8	2.3	131	20	AAV23876
38	8	2.3	131	20	AAV95195
39	8	2.3	162	21	AAV33452
40	8	2.3	165	22	AAV63414
41	8	2.3	219	22	AAV72164
42	8	2.3	231	22	AAV68902
43	8	2.3	231	22	AAV68903
44	8	2.3	231	22	AAV68904
45	8	2.3	231	22	AAV68905

ALIGNMENTS

RESULT 1	AAV45466	standard; Protein; 144 AA.
XX	AC	AAV45466;
XX	AC	AAV45466;
DT	18-OCT-2000	(first entry)
XX	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 57085.
DE	XX	Protein identification assay; signal transduction pathway; metabolic pathway;
XX	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	KW	termination sequence.
XX	XX	Arabidopsis thaliana.
OS	XX	Arabidopsis thaliana.
XX	XX	Arabidopsis thaliana.
PN	XX	Arabidopsis thaliana.
XX	XX	Arabidopsis thaliana.
PD	06-SEP-2000.	
XX	XX	
PF	25-FEB-2000;	2000EP-0301439.
XX	XX	
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.

Protein derived fr
Protein derived fr
P. americanus anti
P. americanus anti
P. americanus anti
P. americanus anti
Winter flounder 11
Winter flounder 11
Meiling pt. depre
P. americanus anti
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Human alpha 2 C4 a
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Winter flounder an
P. americanus anti
Human endometrium
A hypothetical mat
Hypothetical hybr
Zea mays protein f
Human breast cance
Human RNA metaboli
Neisseria meningit
Neisseria meningit
Neisseria meningit
Neisseria meningit

PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139464.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140891.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149930.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151630.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 24-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.98; Score 10; DB 21; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 SALVAAAAA 19
Db 5 salavaaaa 14

RESULT 2

AAAG45465
ID AAG45465 standard; Protein: 209 AA.

AC AAG45465;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 57084.

KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126284.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-015753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158365.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 209;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAA 19
Db 70 salavaaaaa 79

RESULT 3
AAC49854
ID AAC49854 standard; Protein; 268 AA.
XX AAC49854;
AC
AC AAG49854;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63112.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 30-JUN-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160815.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 268;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAAA 19
| | | | | | | | | |
Db 5 salavaaaaa 14

RESULT 4
AAG19454
ID AAG19454 standard; Protein; 319 AA.
XX
AC AAG19454;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21262.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200DEP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154179.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155146.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.
Query Match 2.9%; Score 10; DB 21; Length 319;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 SALAVAAAAA 19
|||||||
Db 5 salavaaaa 14
RESULT 5
AAG49849
ID AAG49849 standard; Protein; 322 AA.
XX
AC AAG49849;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63105.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0127464.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.

PR	09-AUG-1999;	9905-0147935
PR	10-AUG-1999;	9905-0148171
PR	11-AUG-1999;	9905-0148319
PR	12-AUG-1999;	9905-0148341
PR	13-AUG-1999;	9905-0148565
PR	16-AUG-1999;	9905-0149368
PR	17-AUG-1999;	9905-0149175
PR	18-AUG-1999;	9905-0149426
PR	20-AUG-1999;	9905-0149722
PR	20-AUG-1999;	9905-0149723
PR	20-AUG-1999;	9905-0149929
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151086
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151438
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0152363
PR	10-SEP-1999;	9905-0153070
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154039
PR	16-SEP-1999;	9905-0154018
PR	20-SEP-1999;	9905-0154779
PR	22-SEP-1999;	9905-0155139
PR	23-SEP-1999;	9905-0155486
PR	24-SEP-1999;	9905-0155569
PR	28-SEP-1999;	9905-0156458
PR	29-SEP-1999;	9905-0156557
PR	04-OCT-1999;	9905-0157117
PR	05-OCT-1999;	9905-0157753
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158232
PR	12-OCT-1999;	9905-0158369
PR	13-OCT-1999;	9905-0158293
PR	13-OCT-1999;	9905-0159294
PR	14-OCT-1999;	9905-0159325
PR	14-OCT-1999;	9905-0159329
PR	14-OCT-1999;	9905-0159331
PR	14-OCT-1999;	9905-0159331
PR	14-OCT-1999;	9905-0159637
PR	18-OCT-1999;	9905-0159684
PR	21-OCT-1999;	9905-0160741
PR	21-OCT-1999;	9905-0160747
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160777
PR	21-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160981
PR	25-OCT-1999;	9905-0161404
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161406
PR	26-OCT-1999;	9905-0161359
PR	26-OCT-1999;	9905-0161360
PR	26-OCT-1999;	9905-0161361
PR	28-OCT-1999;	9905-0161920
PR	28-OCT-1999;	9905-0161922
PR	29-OCT-1999;	9905-0161993
PR	29-OCT-1999;	9905-0162142

```

Query Match          2.9%; Score 10; DB 21; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0Y 10 SALAAVAAAA 19

```

```
DB      5 |||||||||
        5 salavaaaa 14

RESULT  6
AAG49853
ID      AAG49853 standard; Protein: 333 AA.
XX
AC      AAG49853;
XX
DT      18-OCT-2000 (first entry)
XX
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 63111.
XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX
OS      Arabidopsis thaliana.
XX
PN      EPI033405-A2.
XX
PD      06-SEP-2000.
XX
FE      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128232.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0132048.
PR      30-APR-1999; 99US-0132407.
PR      04-MAY-1999; 99US-0132484.
PR      05-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      06-MAY-1999; 99US-0132487.
PR      07-MAY-1999; 99US-0132863.
PR      11-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      14-MAY-1999; 99US-0134370.
PR      18-MAY-1999; 99US-0134768.
PR      19-MAY-1999; 99US-0134941.
PR      20-MAY-1999; 99US-0135124.
PR      21-MAY-1999; 99US-0135353.
PR      24-MAY-1999; 99US-0135629.
PR      25-MAY-1999; 99US-0136021.
PR      27-MAY-1999; 99US-0136392.
PR      28-MAY-1999; 99US-0136782.
PR      01-JUN-1999; 99US-0137222.
PR      03-JUN-1999; 99US-0137528.
PR      04-JUN-1999; 99US-0137502.
PR      07-JUN-1999; 99US-0137724.
PR      08-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      10-JUN-1999; 99US-0138847.
PR      14-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      16-JUN-1999; 99US-0139453.
PR      17-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.

PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139763.
PR      21-JUN-1999; 99US-0139817.
PR      22-JUN-1999; 99US-0139899.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.
PR      01-JUL-1999; 99US-0141842.
PR      02-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144088.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144325.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      19-JUL-1999; 99US-0144335.
PR      20-JUL-1999; 99US-0144352.
PR      20-JUL-1999; 99US-0144632.
PR      20-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145088.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145218.
PR      23-JUL-1999; 99US-0145224.
PR      26-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
PR      02-AUG-1999; 99US-0146389.
PR      03-AUG-1999; 99US-0147038.
PR      04-AUG-1999; 99US-0147204.
PR      04-AUG-1999; 99US-0147302.
PR      05-AUG-1999; 99US-0147192.
PR      05-AUG-1999; 99US-0147260.
PR      06-AUG-1999; 99US-0147303.
PR      06-AUG-1999; 99US-0147410.
PR      09-AUG-1999; 99US-0147493.
PR      09-AUG-1999; 99US-0147935.
PR      10-AUG-1999; 99US-0148171.
PR      11-AUG-1999; 99US-0148319.
PR      12-AUG-1999; 99US-0148341.
PR      13-AUG-1999; 99US-0148565.
PR      13-AUG-1999; 99US-0148684.
PR      16-AUG-1999; 99US-0149368.
```

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161892.
PR 28-OCT-1999; 99US-0161893.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 333;
Best Local Similarity 100.0%; Pred No. 0.32;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 SALAVAAAAA 19
|||||||
Db 70 salavaaaaa 79

RESULT 7
AAG19453
ID AAG19453 standard; Protein; 384 AA.

XX AC AAG19453;
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21261.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.

```
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159299.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAAA 19
Db 70 salavaaaaa 79

RESULT 8
ID AAG49848 standard; Protein; 387 AA.
XX AAG49848;
AC AAG49848;
XX 18-OCT-2000 (first entry)
DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63104.
DE XX
```

KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134321.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.

PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 02-JUL-1999; 9905-0142154.
PR 06-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148655.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149176.
PR 20-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.

PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 387;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 SALAVAAAAA 19
 |||||
 DB 70 salavaaaa 79

RESULT 9
 AAB72380
 ID AAB72380 standard; protein: 337 AA.

AC AAB72380;

DT 22-MAY-2001 (first entry)

DE Hairy protein amino acid sequence.

KW Cardiac helix-loop-helix factor; CHF; transcription factor;

KW cardiomyocyte; proliferation; myocarditis; myocardial infarction;

KW cardiomyocyte regeneration; angiogenesis inhibitor; differentiation;

KW smooth muscle cell growth; heart disease; hairy.

XX Drosophila sp.

OS
 XX

PN WO200112126-A2.
 XX 22-FEB-2001.
 XX 11-AUG-2000; 2000WO-US21858.
 XX 13-AUG-1999; 99US-0148974.
 XX (HARD) HARVARD COLLEGE.
 PA (LEEM/) LEE M.
 XX Lee M, Chin MT;
 XX WPI: 2001-211119/21.
 DR
 XX
 PT Novel cardiovascular helix-loop-helix factor polypeptide and
 PT polynucleotide useful for regenerating heart tissue and promoting
 PT smooth muscle differentiation to treat heart disease or heart injury -
 XX Disclosure: Fig 1; 46pp; English.

This invention relates to human cardiovascular helix-loop-helix factor (CHF) polypeptides. CHF is a transcription factor expressed in cardiovascular tissue, which contains a basic helix-loop-helix domain. CHF-1 shares structural homology with hairy, a Drosophila protein essential for the development of the peripheral nervous system. The CHF polypeptide is useful for expanding an endogenous population of proliferative cardiomyocytes in a heart tissue, promoting proliferation of a cardiomyocyte in a mammal having myocarditis or that which has suffered myocardial infarction and for regenerating cardiomyocytes in vivo. DNA encoding a CHF-1 polypeptide is also useful for promoting proliferation of a cardiomyocyte, regenerating cardiomyocytes in vitro, inhibiting angiogenesis in a tissue, inducing differentiation of smooth muscle cells in a mammalian tissue, preferably venous tissue, inducing growth of smooth muscle cells in a vein explant, promoting smooth muscle cell regeneration in an injured or diseased vascular tissue (venous or arterial) and for reducing vein graft stenosis in a mammal by contacting the tissue ex vivo with the DNA prior to implantation of the tissue into an artery of the mammal. The present sequence represents the hairy protein from Drosophila, with which CHF-1 shares homology.

Sequence 337 AA;

Query Match 2.6%; Score 9; DB 22; Length 337;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
 |||||
 DB 238 lavaaaaa 246

RESULT 10
 AAM27490
 ID AAM27490 standard; peptide: 33 AA.

AC AAM27490;

DT 20-APR-1998 (first entry)

DE Arctic fish antifreeze peptide.

KW Antifreeze protein; cryoprotectant; arctic fish; transfusion;

KW haematopoietic cell; freeze-drying.

XX Pisces.

OS DE19611969-A1.

PN 02-OCT-1997.

PD 26-MAR-1996; 96DE-1011969.

XX PF

XX 26-MAR-1996; 96DE-1011969.
 XX (RADU/) RADULESCU R T.
 XX
 XX Radulescu RT;
 XX WPI; 1997-481546/45.
 XX
 XX Use of antifreeze peptide in transfusion medicine - especially as
 PT cryoprotectant for hematopoietic cells
 XX
 XX Disclosure; Column 1; 2pp; German.
 XX
 XX This sequence represents an antifreeze peptide which can be used in the
 CC field of transfusion medicine. This peptide can be used as a
 CC cryoprotectant to protect hematopoietic cells during freezing, thawing
 CC and freeze-drying. The peptide has the same activity as the native
 CC arctic fish protein from which it is derived and should have no
 CC significant side effects as it rapidly degrades in vivo.
 XX
 XX Sequence 33 AA;

Query Match 2.3%; Score 8; DB 18; Length 33;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAALT 22
 DB 2 aaaaalt 9

RESULT 11

ID AAR26105 standard; Protein; 37 AA.
 AC AAR26105;
 DT 01-FEB-1993 (first entry)
 DE Antifreeze protein from Winter Flounder.
 XX
 XX Thermal hysteresis protein; Ice crystal.
 XX
 OS Pseudopleuronectes americanus.
 PN WO9212722-A.
 PD 06-AUG-1992.
 PF 17-JAN-1992; 92WO-US00452.
 PR 17-JAN-1991; 91WO-US00351.
 PA (REGC) UNIV CALIFORNIA.
 PI Devries AL, Rubinsky B;
 DR WPI; 1992-284413/34.
 PT Compn. contg. thermal hysteresis protein - used to protect and
 PT preserve biological materials in non-physiological temps. and
 PT conditions
 PS Claim 28; Page 21; 126pp; English.
 CC The antifreeze protein from the Winter Flounder lacks sugars (C.f.
 CC the antifreeze glycopeptides of Nototheniidae fish, e.g. AAR26104)
 CC and instead has high percentages of hydrophilic amino acids (Thr
 CC and Asp) while still retaining a large amount of Ala. The AFP from
 CC P.americanus is one of the preferred thermal hysteresis proteins for
 CC use in the compositions of the invention.

CC See also AAR26106.
 XX
 XX Sequence 37 AA;
 SO

Query Match 2.3%; Score 8; DB 13; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAALT 22
 DB 6 aaaaalt 13

RESULT 12

AAY23879
 ID AAY23879 standard; Protein; 37 AA.
 AC AAY23879;
 DT 06-OCT-1999 (first entry)
 DE Protein derived from winter flounder antifreeze protein (AFP) A.
 XX
 XX Antifreeze protein; AFP; Ice crystal; Winter flounder;
 XX Pseudopleuronectes americanus; Ice crystal growth; frozen product;
 KW ice cream; paint; transplant organ; yeast culture; brewing; baking;
 KW frozen.
 XX
 XX Synthetic.
 OS Pseudopleuronectes americanus.
 PN US5928877-A.
 PD 27-JUL-1999.
 PF 20-NOV-1997; 97US-0975166.
 PR 20-JUL-1992; 92US-0917216.
 PR 19-SEP-1989; 89US-0409217.
 PR 28-FEB-1990; 90US-0486333.
 PR 12-JAN-1994; 94US-0180524.
 PR 20-NOV-1997; 97US-0975166.
 PA (MILL-) MILLER BREWING CO.
 PI Cronan CL, Lusk L;
 PI WPI; 1999-429496/36.
 DR
 DR
 DR
 PT
 PT
 PT
 PS
 PS
 PS

Detecting the presence of antifreeze proteins in samples
 Example 1; Column 14; 24pp; English.

The present sequence is derived from an antifreeze protein (AFP) A. The
 CC specification describes a method for detecting the presence of
 CC functional AFPs in samples. The method comprises freezing a test
 CC solution which contains sucrose, has a melting point (mp), and in
 CC which ice crystals are formed, raising the temperature of the test
 CC solution to its mp to partially melt the ice crystals, lowering the
 CC temperature 1 to 2 degrees Celsius below the mp, which causes the
 CC partially melted crystals to begin to grow, and determining the
 CC presence of functional AFPs by observing the sample for the growth
 CC of hexagonal ice crystals. The method may be used for detecting the
 CC presence of functional AFPs in samples. AFPs (which may be isolated
 CC from fish in cold polar marine waters such as the winter flounder
 CC (Pseudopleuronectes americanus) bind to, and inhibit the formation
 CC of ice crystals in water as it is frozen. They may be used to
 CC prevent ice crystal growth in a wide range of frozen products such
 CC as ice cream and paint. In particular they may be applied to living
 CC cells and tissues (e.g. transplant organs and yeast cultures for
 CC brewing and baking), allowing them to be frozen without the risk of
 CC ice crystal formation, which disrupts cellular membranes and kills

CC the cells making them useless.
 XX
 SQ Sequence 37 AA:

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 15 AAAAAALT 22
 |||||
 Db 6 aaaaalt 13

RESULT 13
 AAY23880
 ID AAY23880 standard; Protein; 37 AA.
 XX
 AC AAY23880;
 XX
 DT 06-OCT-1999 (first entry)
 XX
 DE Protein derived from winter flounder antifreeze protein (AFP) A.
 XX
 KW Antifreeze protein; AFP; ice crystal; winter flounder;
 KW pseudopleuronectus americanus; ice crystal growth; frozen product;
 KW ice cream; paint; transplant organ; yeast culture; brewing; baking;
 KW frozen.
 XX
 OS Synthetic.
 OS Pseudopleuronectus americanus.
 XX
 PN US5928877-A.
 XX
 PD 27-JUL-1999.
 XX
 PF 20-NOV-1997; 97US-0975166.
 XX
 PR 20-JUL-1992; 92US-0917216.
 PR 19-SEP-1989; 89US-0409217.
 PR 28-FEB-1990; 90US-0486333.
 PR 12-JAN-1994; 94US-0180524.
 PR 20-NOV-1997; 97US-0975166.
 XX
 PA (MILL-) MILLER BREWING CO.
 XX
 PI Cronan CL, Lusk L;
 XX
 DR WPI; 1999-429496/36.
 XX
 PT
 XX
 PS
 XX Example 1; Column 14; 24pp; English.
 CC The present sequence is derived from an antifreeze protein (AFP) A. The
 CC specification describes a method for detecting the presence of
 CC functional AFPs in samples. The method comprises freezing a test
 CC solution which contains sucrose, has a melting point (mp), and in
 CC which ice crystals are formed, raising the temperature of the test
 CC solution to its mp to partially melt the ice crystals, lowering the
 CC temperature 1 to 2 degrees Celsius below the mp, which causes the
 CC partially melted crystals to begin to grow, and determining the
 CC presence of functional AFPs by observing the sample for the growth
 CC of hexagonal ice crystals. The method may be used for detecting the
 CC presence of functional AFPs in samples. AFPs (which may be isolated
 CC from fish in cold polar marine waters such as the winter flounder
 CC (Pseudopleuronectus americanus) bind to, and inhibit the formation
 CC of ice crystals in water as it is frozen. They may be used to
 CC prevent ice crystal growth in a wide range of frozen products such
 CC as ice cream and paint. In particular they may be applied to living
 CC cells and tissues (e.g. transplant organs and yeast cultures for
 CC brewing and baking), allowing them to be frozen without the risk of
 CC ice crystal formation, which disrupts cellular membranes and kills

CC the cells making them useless.
 XX
 SQ Sequence 37 AA:

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 15 AAAAAALT 22
 |||||
 Db 6 aaaaalt 13

RESULT 14
 AAM86157
 ID AAM86157 standard; protein; 37 AA.
 XX
 AC AAM86157;
 XX
 DT 11-MAR-1999 (first entry)
 XX
 DE P. americanus antifreeze protein A variant AFP A(10).
 XX
 KW Winter flounder; AFP A; antifreeze protein; transformation; yeast;
 KW Saccharomyces cerevisiae; variant; AFP; ice crystal; food; beverage.
 XX
 OS Pseudopleuronectus americanus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 23
 FT /label=L23D
 FT /note="wild-type Leu is replaced by Asp"
 XX
 XX
 PN US5849537-A.
 XX
 PD 15-DEC-1998.
 XX
 PR 12-JAN-1994; 94US-0180524.
 XX
 PR 20-JUL-1992; 92US-0917216.
 PR 19-SEP-1989; 89US-0409217.
 PR 28-FEB-1990; 90US-0486333.
 PR 12-JAN-1994; 94US-0180524.
 XX
 PA (MILL-) MILLER BREWING CO.
 XX
 PI Barney MC, Bower PA, Chicoye E, Cronan CL, Hulge N;
 PI Kot E, Lusk L, Rhodes T, Tripp M;
 XX
 DR WPI; 1999-094410/08.
 XX
 PT
 XX
 PS
 XX Claim 2; Column 14; 25pp; English.
 CC Vector for expressing antifreeze protein in yeast - for producing
 CC variant of Pseudopleuronectus americanus antifreeze protein

CC The present sequence represents a variant of the antifreeze protein A
 CC (AFP A). The invention relates to an expression vector for transforming
 CC yeast (Saccharomyces cerevisiae) so that it expresses and secretes an
 CC antifreeze protein with no N-terminal methionine. The vector contains a
 CC DNA sequence encoding a variant of the winter flounder AFP A
 CC (Pseudopleuronectus americanus antifreeze protein A) sequence in which
 CC one amino acid is replaced with a different amino acid. The antifreeze
 CC protein can be used to reduce the rate of ice crystal growth in food,
 CC beverage and nonfood applications.
 XX
 SQ Sequence 37 AA:

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 AAAAALT 22
 |||||
 Db 6 aaaaalt 13

Search completed: August 23, 2001, 11:17:24
 Job time: 196 sec

RESULT 15
 AAW86158
 ID AAW86158 standard; protein; 37 AA.
 AC AAW86158;
 XX
 DT 11-MAR-1999 (first entry)
 XX
 DE P. americanus antifreeze protein A variant AFP A(38).
 XX
 KM Winter flounder; AFP A; antifreeze protein; transformation; yeast;
 XX Saccharomyces cerevisiae; variant; AFP; ice crystal; food; beverage.
 OS Pseudopleuronectes americanus.
 XX Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 36 /label= A36V
 FT /note= "wild-type Ala is replaced by Val"
 PN US5849537-A.
 PD 15-DEC-1998.
 PF 12-JAN-1994; 94US-0180524.
 XX
 PR 20-JUL-1992; 92US-0917216.
 PR 19-SEP-1989; 89US-0409217.
 PR 28-FEB-1990; 90US-0486333.
 PR 12-JAN-1994; 94US-0180524.
 XX
 PA (MILL-) MILLER BREWING CO.
 PI Barney MC, Bower PA, Chlocoy E, Cronan CL, Huige N;
 PI Kot E, Lusk L, Rhodes T, Tripp M;
 XX
 DR MPI; 1999-094410/08.
 DR N-PSDB; AAV84680.
 XX
 PT Vector for expressing antifreeze protein in yeast - for producing
 PT variant of Pseudopleuronectes americanus antifreeze protein
 XX
 PS Example 1; Column 14; 25pp; English.
 XX
 CC The present sequence represents a variant of the antifreeze protein A
 CC (AFP A). The invention relates to an expression vector for transforming
 CC yeast (Saccharomyces cerevisiae) so that it expresses and secretes an
 CC antifreeze protein with no N-terminal methionine. The vector contains a
 CC DNA sequence encoding a variant of the winter flounder AFP A
 CC (Pseudopleuronectes americanus antifreeze protein A) sequence in which
 CC one amino acid is replaced with a different amino acid. The antifreeze
 CC protein can be used to reduce the rate of ice crystal growth in food,
 CC beverage and nonfood applications.
 CC
 SQ Sequence 37 AA;

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 AAAAALT 22
 |||||
 Db 6 aaaaalt 13

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:18:03 ; Search time 32.82 Seconds
(without alignments)
810.022 Million cell updates/sec

Title: US-09-654-652a-3
Perfect score: 349
Sequence: 1 MNIKKTAVKASALAVAAAAA.....AKGAKVNPNGHKRYVNFEEH 349

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	349	100.0	349	2 A44507	licheninase (EC 3.1.1.1)
2	349	100.0	349	2 A44507	licheninase (EC 3.1.1.1)
3	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
4	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
5	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
6	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
7	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
8	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
9	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
10	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
11	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
12	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
13	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
14	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
15	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
16	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
17	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
18	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
19	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
20	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
21	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
22	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
23	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
24	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
25	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
26	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
27	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
28	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)
29	348	100.0	348	2 A44507	licheninase (EC 3.1.1.1)

30	8	2.3	317	2 C70874	hypothetical prote
31	8	2.3	339	2 T26328	hypothetical prote
32	8	2.3	368	2 G83463	probable methyltra
33	8	2.3	368	2 T46615	chemotaxis protein
34	8	2.3	374	2 T03875	probable homeobox
35	8	2.3	375	2 T03874	probable homeobox
36	8	2.3	378	2 T06512	DNA-binding protei
37	8	2.3	381	2 S29560	fructose-bisphosph
38	8	2.3	392	2 B48423	homeoic protein e
39	8	2.3	401	2 A48423	engrailed homeod
40	8	2.3	407	2 T06408	probable fructose-
41	8	2.3	414	2 A86229	hypothetical prote
42	8	2.3	424	2 S09884	hypothetical prote
43	8	2.3	429	2 T07815	S-luciferase
44	8	2.3	429	2 D07884	probable Rieske [2
45	8	2.3	459	2 S13064	Id-myo-inositol-tr

ALIGNMENTS

RESULT 1
A44507
licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Reuther, R.M.; Erfle, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1.3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-references: EMBL:M33676; NID:q148575; PIDN:AAA24896.1; PID:q148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match	100.0%	Score 349	DB 2	Length 349
Best Local Similarity	100.0%	Pred. No. 0		
Matches 349	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1 MNIKKTAVKASALAVAAAAAALTTNWSAKDFSGAEYTLLEVOYGRFEARMKAAASGYVS 60			
DB	1 MNIKKTAVKASALAVAAAAAALTTNWSAKDFSGAEYTLLEVOYGRFEARMKAAASGYVS 60			
QY	61 SMFLYQNGSEIADGRPWVEVDIEVLGKNPQSFQSNITGKAGAQRTSEKHNAVSPADQA 120			
DB	61 SMFLYQNGSEIADGRPWVEVDIEVLGKNPQSFQSNITGKAGAQRTSEKHNAVSPADQA 120			
QY	121 PFTYGLWPTNRYVNTVQGEVYKTEGGQVSNLTGTGIRFNLMSSESAVAVGQDESKL 180			
DB	121 PFTYGLWPTNRYVNTVQGEVYKTEGGQVSNLTGTGIRFNLMSSESAVAVGQDESKL 180			
QY	181 PFTYGLWPTNRYVNTVQGEVYKTEGGQVSNLTGTGIRFNLMSSESAVAVGQDESKL 240			
DB	181 PFTYGLWPTNRYVNTVQGEVYKTEGGQVSNLTGTGIRFNLMSSESAVAVGQDESKL 240			
QY	241 YSRDMLILALTRKQSEFNGVPRDDEPAPOSSSSAPASSSSVPASSSSSAFV 300			
DB	241 YSRDMLILALTRKQSEFNGVPRDDEPAPOSSSSAPASSSSVPASSSSSAFV 300			
QY	301 PSSSSATNAIHGMRTTPAVAKENHNLVNAKAKVNPNGHKRYVNFEEH 349			
DB	301 PSSSSATNAIHGMRTTPAVAKENHNLVNAKAKVNPNGHKRYVNFEEH 349			
RESULT 2				
T52635	mitogen-activated protein kinase kinase (EC 2.7.1.1) alpha [imported] - Arabidopsis t			
C:Species:	Arabidopsis thaliana (mouse-ear cress)			
C:Date:	24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000			

C:Accession: T52635
 R:Hama, A.; Jouanvic, S.; Leprince, S.; Kreis, M.; Henry, Y.
 Plant Sci. 140, 41-52, 1999
 A:Title: Molecular characterization and expression of an Arabidopsis thaliana L. MAP kin
 A:Reference number: 226147
 A:Accession: T52635
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-348 <HAM>
 A:Cross-references: EMBL:Y07694; PIDN:CAA68958.1
 A:Experimental source: cultivar Columbia; seedling
 C:Genetics:
 A:Gene: MAP2Kalpha
 A:Map position: 3
 C:Superfamily: kinase-related transforming protein; protein kinase homology
 C:Keywords: ATP; phosphotransferase

Query Match 2.9%; Score 10; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 0.2;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 273 SSSSAPASSS 282
 Db 46 SSSSAPASSS 55
 |||||||||

RESULT 3
 T51340
 mitogen-activated protein kinase kinase (EC 2.7.1.-) 5 [imported] - Arabidopsis thaliana
 N:Alternate names: MAP kinase kinase 5
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 08-Sep-2000
 C:Accession: T51340
 R:Richmura, K.; Mizoguchi, T.; Hayashida, N.; Seki, M.; Shinzaki, K.
 DNA Res. 5, 341-348, 1998
 A:Title: Molecular cloning and characterization of three cDNAs encoding putative mitogen
 A:Reference number: 225272; MUID:99156228
 A:Accession: T51340
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-348 <ICH>
 A:Cross-references: EMBL:AB015316; PIDN:BAA28831.1
 C:Genetics:
 A:Gene: ATKK5
 C:Function:
 A:Description: (EC 2.7.1.-): mitogen-activated protein kinase [validated; MUID:99
 C:Superfamily: kinase-related transforming protein; protein kinase homology
 C:Keywords: phosphotransferase; protein kinase

Query Match 2.9%; Score 10; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 0.2;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 273 SSSSAPASSS 282
 Db 46 SSSSAPASSS 55
 |||||||||

RESULT 4
 S59632
 endo-1,4-beta-xylosanase (EC 3.2.1.8) B precursor - Cellvibrio mixtus
 C:Species: Cellvibrio mixtus
 C:Date: 15-Feb-1996 #sequence_revision 11-Apr-1997 #text_change 18-Jun-1999
 C:Accession: S59632; S52742
 R:Millward-Sadler, S.J.; Davidson, K.; Hazlewood, G.P.; Black, G.W.; Gilbert, H.J.; Cla
 Biochem. J. 312, 39-48, 1995
 A:Title: Novel cellulose-binding domains, NodB homologues and conserved modular architec
 A:Reference number: S59631; MUID:96077124
 A:Accession: S59632
 A:Molecule type: DNA

A:Residues: 1-621 <MIL>
 A:Cross-references: EMBL:Z48926; NID:g757808; PIDN:CAA8762.1; PID:g757809
 C:Genetics:
 A:Gene: xynB
 C:Function:
 A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
 A:Pathway: xylan degradation
 C:Superfamily: Pseudomonas endo-1,4-beta-xylosanase F; Streptomyces endo-1,4-beta-xylan
 F:1-19/Domains: glycosidase; hydrolase; polysaccharide degradation
 F:20-621/Product: endo-1,4-beta-xylosanase B #status predicted <MAT>
 F:302-615/Domains: Streptomyces endo-1,4-beta-xylosanase A homology <SXY>
 F:403,516/Active site: Glu #status predicted

Query Match 2.9%; Score 10; DB 1; Length 621;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 274 SSSAPASSS 283
 Db 89 SSSAPASSS 98
 |||||||||

RESULT 5
 B84259
 hypothetical protein Vng1026h [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: B84259
 R:N, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
 ; Leitauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: AB4160; MUID:20504483
 A:Accession: B84259
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-184 <STO>
 A:Cross-references: GB:AE004437; NID:g10580580; PIDN:AG19438.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG1026H

Query Match 2.6%; Score 9; DB 2; Length 184;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
 Db 137 LAVAAAAA 145
 |||||||||

RESULT 6
 S06956
 segmentation protein hairy - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
 C:Accession: S06956; S06700
 R:Rushlow, C.A.; Hogan, A.; Pinchin, S.M.; Howe, K.M.; Lardelli, M.; Ish-Horowicz, D.
 EMBO J. 8, 3095-3103, 1989
 A:Title: The Drosophila hairy protein acts in both segmentation and bristle pattern
 A:Reference number: S06956; MUID:90059896
 A:Accession: S06956
 A:Molecule type: DNA
 A:Residues: 1-337 <RUS>
 A:Cross-references: GB:X15904; GB:S63792; GB:X16632; NID:g8048; PIDN:CAA34018.1; PID:
 A:Note: 292-ser was also found
 C:Genetics:
 A:Gene: hairy
 A:Cross-references: FlyBase:FBgn0001168

A:Map position: 66D
A:introns: 33/3; 65/3
C:Keywords: DNA binding; transcription regulation

Query Match 2.6%; Score 9; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 12 LAVAAAAA 20
| | | | | | | | | |
Db 238 LAVAAAAA 246

RESULT 7
A:72599
hypothetical protein APE1257 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72599
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6: 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Cremonarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339
A:Accession: A72599
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-439 <KAM>
A:Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BA80247.1; PID:d1044033; PID:g5104821
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1257

Query Match 2.6%; Score 9; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 SALAVAAA 18
| | | | | | | | | |
Db 398 SALAVAAA 406

RESULT 8
A:A6600
methylmalonate-semialdehyde dehydrogenase (acylating) (EC 1.2.1.27) precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-May-2000
C:Accession: A46600
R:Deichalte, I.; Berthiaume, L.; Pesce, S.M.; Patton, W.F.; Resh, M.D.
J. Biol. Chem. 268, 13738-13747, 1993
A:Title: Novel use of an Iodo-myristyl-CoA analog identifies a semialdehyde dehydrogenase
A:Reference number: A46600; MUID:93293905
A:Accession: A46600
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-537 <DEI>
A:Cross-references: GB:L08643; NID:g289441; PIDN:AAA30650.1; PID:g289442
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: coenzyme A; mitochondrion; oxidoreductase
F:80-336/Domain: aldehyde dehydrogenase homology <ALDD>
F:319/Active site: Cys #status predicted

Query Match 2.6%; Score 9; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 AVAAAAAL 21
| | | | | | | | | |
Db 5 AVAAAAAL 13

RESULT 9
A:729064
hyaluronate lyase homolog - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: J29064
R:Redenbach, M.; Kleiser, H.M.; Denapalte, D.; Eichner, A.; Cullum, J.; Kinashi, H.; H Mol. Microbiol. 21, 77-96, 1996
A:Title: A set of ordered cosmid and a detailed genetic and physical map for the 8 M
A:Reference number: Z20556; MUID:97000351
A:Accession: J29064
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-776 <RED>
A:Cross-references: EMBL:AL031124; NID:e1312893; PID:e1312908; PIDN:CA19982.1
C:Genetics:
A:Note: SCIC2.15

Query Match 2.6%; Score 9; DB 2; Length 776;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 15 AAAAAALTT 23
| | | | | | | | | |
Db 18 AAAAAALTT 26

RESULT 10
A:148719
protein kinase C (EC 2.7.1.1) mu precursor - mouse
M:Alternate names: protein kinase D
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 11-Jun-1999
C:Accession: I48719
R:Valverde, A.M.; Simmett-Smith, J.; Van Lint, J.; Rozenfurt, E.
Proc. Natl. Acad. Sci. U.S.A. 91, 8572-8576, 1994
A:Title: Molecular cloning and characterization of protein kinase D: a target for dia
A:Reference number: I48719; MUID:94359973
A:Accession: I48719
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-918 <RES>
A:Cross-references: EMBL:Z34524; NID:g520877; PIDN:CA84283.1; PID:g520878
C:Function:
A:Description: catalyzes the formation of peptidyl-serine-phosphate or peptidyl-threo
A:Note: expressed at low levels in a variety of tissues; phosphorylates the peptide s
sely related human enzyme, this protein is reported to bind phorbol esters
C:Superfamily: protein kinase C mu; protein kinase C zinc-binding repeat homology; pr
C:Keywords: ATP; autophosphorylation; duplication; phorbol ester binding; phospholip
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-918/Product: protein kinase mu #status predicted <MAT>
F:145-194/Domain: protein kinase C zinc-binding repeat homology <KZ1>
F:277-326/Domain: protein kinase C zinc-binding repeat homology <KZ2>
F:587-845/Domain: protein kinase C zinc-binding repeat homology <KZ3>
F:595-603/Region: protein kinase homology <KIN>
F:618,636,712,714/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 2.6%; Score 9; DB 1; Length 918;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 AVAAAAAL 21
| | | | | | | | | |
Db 19 AVAAAAAL 27

RESULT 11
A:S42718
nuclear pore complex protein nup153 - human
C:Species: Homo sapiens (man)

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C:Accession: S42718; S37477
R:McMorrow, I.; Bastos, R.; Horton, H.; Burke, B.
Biochim. Biophys. Acta 1217, 219-223, 1994
A:Title: Sequence analysis of a cDNA encoding a human nuclear pore complex protein, hnnpf
A:Reference number: S42718; MUID:94154002
A:Molecule type: mRNA
A:Residues: 1-1475 <MCNC>
A:Cross-references: EMBL:Z25535; NID:g406224; PIDN:CAA80982.1; PID:g406225

Query Match 2.3%; Score 8; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 282 SSVASSSS 290
|||||||
DB 823 SSVASSSS 831

RESULT 12

FDPL3W
antifreeze protein 3 - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C>Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 23-Aug-1996
C:Accession: A03192
R:Devries, A.L.; Lito, Y.
Biochim. Biophys. Acta 495, 388-392, 1977
A:Title: Structure of a peptide antifreeze and mechanism of adsorption to ice.
A:Reference number: A03192; MUID:78060969
A:Accession: A03192
A:Molecule type: protein
A:Residues: 1-37 <DEW>
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 2.3%; Score 8; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAAALT 22
|||||||
DB 6 AAAAAALT 13

RESULT 13

PN0589
tyrosine 3-monooxygenase (EC 1.14.16.2) type 1 and 2 - chimpanzee (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Pan troglodytes (chimpanzee)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0589
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398
A:Accession: PN0589
A:Molecule type: genomic RNA
A:Residues: 1-45 <ICH>
A:Cross-references: GB:L14791
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis.
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; monooxygenase; oxidoreductase

Query Match 2.3%; Score 8; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAA 20

DB 22 AVAAAAAA 29
|||||||

RESULT 14

PN0590
tyrosine 3-monooxygenase (EC 1.14.16.2) type 1 and 2 - gorilla (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Gorilla gorilla (gorilla)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0590
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398
A:Accession: PN0590
A:Molecule type: genomic RNA
A:Residues: 1-45 <ICH>
A:Cross-references: GB:L14797
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis.
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; monooxygenase; oxidoreductase

Query Match 2.3%; Score 8; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAA 20
|||||||
DB 22 AVAAAAAA 29

RESULT 15

PN0591
tyrosine 3-monooxygenase (EC 1.14.16.2) type 1 and 2 - orangutan (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Pongo pygmaeus (orangutan)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0591
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398
A:Accession: PN0591
A:Molecule type: genomic RNA
A:Residues: 1-45 <ICH>
A:Cross-references: GB:L14800
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis.
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; monooxygenase; oxidoreductase

Query Match 2.3%; Score 8; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAA 20
|||||||
DB 22 AVAAAAAA 29

Search completed: August 23, 2001, 11:18:05
Job time: 212 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:18:28 ; Search time: 19.47 Seconds

(without alignments)
469.759 Million cell updates/sec

Title: US-09-654-652a-2

Perfect score: 267
Sequence: 1 MVSARDFSGAEIYLEVOY.....PNSSVDKLAALHHHHH 267

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size: 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	92.5	349	1 GUB_FIBSU	P17989 fibroblast
2	8	3.0	317	1 CYSD_RHIME	P56892 rhizobium m
3	8	3.0	407	1 F16P_PEA	P46275 plasm saliv
4	8	3.0	429	1 OCRA_MCTU	Q10387 mycobacteri
5	7	2.6	59	1 HPN_HELPY	Q48251 mycobacte
6	7	2.6	95	1 YWKE_BACSU	P45874 bacillus su
7	7	2.6	119	1 TGH2_YEAST	P53161 saccharomyc
8	7	2.6	301	1 NODP_RHIS3	P72338 rhizobium s
9	7	2.6	345	1 HEMZ_BRAJA	P28602 bradyrhizob
10	7	2.6	351	1 HRPX_PLALO	P04929 plasmodium
11	7	2.6	369	1 MIX2_XENLA	Q91685 xenopus lae
12	7	2.6	377	1 MIX1_XENLA	P21711 xenopus lae
13	7	2.6	396	1 HMPA_ECOLI	P42432 escherichia
14	7	2.6	440	1 YDUN_HABIN	P45079 haemophilus
15	7	2.6	468	1 ETS2_MOUSE	P15037 mus musculu
16	7	2.6	469	1 ETS2_HUMAN	P15036 homo sapien
17	7	2.6	473	1 KNOB_PLAFA	P13817 plasmodium
18	7	2.6	484	1 MEC2_MOUSE	Q92246 mus musculu
19	7	2.6	486	1 MEC2_HUMAN	P51608 homo sapien
20	7	2.6	492	1 MEC2_RAT	Q00566 rattus norv
21	7	2.6	495	1 BRN1_MOUSE	P31361 mus musculu
22	7	2.6	497	1 BRN1_RAT	Q63262 rattus norv
23	7	2.6	500	1 BRN1_HUMAN	P20264 homo sapien
24	7	2.6	546	1 BGAM_HUMAN	P16279 homo sapien
25	7	2.6	611	1 SNF1_CANGA	Q00372 candida gla
26	7	2.6	634	1 KNOB_PLAFA	P09346 plasmodium
27	7	2.6	657	1 KNOB_PLAFA	P06719 plasmodium
28	7	2.6	665	1 LAMA_XENLA	P11048 xenopus lae
29	7	2.6	677	1 BGAL_HUMAN	P16278 homo sapien
30	7	2.6	994	1 MISA_DICDI	P22467 dictyosteli
31	7	2.6	1006	1 RAT1_YEAST	Q02792 saccharomyc
32	7	2.6	1024	1 SYIP_STAUF	P41368 staphylococ
33	7	2.6	1025	1 SLAP_CAUCR	P35828 caulobacter

34	6	2.2	16	1 LPH1_ECOLI	P03058 escherichia
35	6	2.2	88	1 WIRA_WHEAT	Q01482 triticum ae
36	6	2.2	95	1 WHEB_BPFI1	O80294 bacterioph
37	6	2.2	114	1 ASR2_LYCES	P37219 lycopersico
38	6	2.2	131	1 YIT1_STAUF	P41370 staphylococ
39	6	2.2	132	1 FOLB_MYCLE	O69529 mycobacteri
40	6	2.2	133	1 Y05C_BP74	P39258 bacterioph
41	6	2.2	137	1 CYB5_ORYSA	P49100 oryza sativ
42	6	2.2	140	1 PRO3_WHEAT	P49234 triticum ae
43	6	2.2	141	1 PRO2_WHEAT	P49233 triticum ae
44	6	2.2	143	1 Y970_MERJA	O58380 methanococ
45	6	2.2	146	1 CYTC_HUMAN	P01034 homo sapien

ALIGNMENTS

RESULT	ID	GUB_FIBSU	STANDARD	PRT	349 AA.
AC	P17989				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)				
DE	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).				
OS	Fibroblast succinogenes (Bacteroides succinogenes).				
OC	Bacteria; Fibroblast/Actinobacteria group; Fibroblast group;				
CC	Fibroblast.				
OX	NCBI_TaxID=833;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.				
RC	STRAIN-ISOLATE S85;				
RX	MEDLINE=90299807; PubMed=2193918;				
RA	Teather R.M., Ertle J.D.;				
RT	*DNA sequence of a fibroblast succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.;				
RL	J. Bacteriol. 172:3837-3841(1990).				
CC	-I- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.				
CC	-I- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M33676; AAA24896.1; -.				
DR	PIR; A44507; A44507.				
DR	HSSP; P23904; IAKK.				
DR	InterPro; IPR000757; -.				
DR	Pfam; PF00722; Glyco_hydro.16; 1.				
DR	PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.				
KW	Hydrolyase; Glycosidase; Signal; Repeat.				
FT	SIGNAL	1	27		
FT	CHAIN	28	349		
FT	ACT_SITE	79	79		
FT	ACT_SITE	83	83		
FT	DOMAIN	271	307		
FT	REPEAT	271	277		
FT	REPEAT	278	284		
FT	REPEAT	285	291		
FT	REPEAT	292	298		
FT	REPEAT	301	307		
SO	SEQUENCE	349 AA;	37737 MM;	16DCAF5BDEFC578A CRC64;	

Query Match 92.5%; Score 247; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 1,9e+244;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VSKADSGALTYLLEEVYGFKFEKRMKMAAAGTVMFLYONGSEFLADGRPVENVIEV 61
 DB 25 VSKADSGALTYLLEEVYGFKFEKRMKMAAAGTVMFLYONGSEFLADGRPVENVIEV 84

QY 62 LGRNPGSFQSNITITGKAGAKTSEKHHAVSPADQAFHTYGLTEPNTVYVMTVDGGEVRK 121
 DB 85 LGRNPGSFQSNITITGKAGAKTSEKHHAVSPADQAFHTYGLTEPNTVYVMTVDGGEVRK 144

QY 122 TEGGQVSNLTGTGGLRNLMSSSAAMVGOFPDSKLPFPQIMWKYKKTTCGCGEGSD 181
 DB 145 TEGGQVSNLTGTGGLRNLMSSSAAMVGOFPDSKLPFPQIMWKYKKTTCGCGEGSD 204

QY 182 FTLDWNTDFTFGSRKSGDWTGDNVDTLTKNITSRDGMLTLATRKGGSPFNGQVP 241
 DB 205 FTLDWNTDFTFGSRKSGDWTGDNVDTLTKNITSRDGMLTLATRKGGSPFNGQVP 264

QY 242 RDDEPAP 248
 DB 265 RDDEPAP 271

RESULT 2
 CYS_D_RHIME STANDARD; PRT; 317 AA.

AC P56892;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SULFATE ADENYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLATE TRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT).
 GN CYS_D.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=99395034; PubMed=10464198;
 RA Abola A.P., Willis M.G., Wang R.C., Long S.R.;
 RT "Reduction of adenosine-5'-phosphosulfate instead of 3'-phosphodenosine-5'-phosphosulfate in cysteine biosynthesis by Rhizobium meliloti and other members of the family Rhizobiaceae.";
 RT J. Bacteriol. 181:5280-5287(1999).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
 RA Bolstad P., Gouzy J., Kahn D., Thebaud P., Goffeau A.,
 RA Purrelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
 RA Vandenbol M., Puehler A., Becker A., Weidner S.;
 RL Submitted (May-2000) to the SWISS-PROT data bank.
 CC -1- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE + ADENYLSULFATE.
 CC -1- PATHWAY: FIRST STEP IN THE SULFATE ACTIVATION PATHWAY. THIS REACTION OCCURS EARLY IN THE REDUCTIVE BRANCH OF THE CYSTEINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF CYS_D, THE SMALLER SUBUNIT, AND CYSN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYS_D SUBFAMILY.
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYS_D SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AF158023; AAD55760.1; -
 DR InterPro: IPR002500; -

DR Pfam: PF01507; PAPS_reduct; 1.
 KM Cysteine biosynthesis; Transferase; Nucleotidyltransferase.
 FT CONFLICT 260
 SQ SEQUENCE 317 AA; 36429 MM; 093F2724D22841CB CRC64;

Query Match
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RDGMILTA 227
 DB 234 RDGMILTA 241

RESULT 3
 F16P_PEA STANDARD; PRT; 407 AA.

AC P46275; Q37263;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FRUCTOSE-1,6-BISPHOSPHATASE, CHLOROPLAST PRECURSOR (EC 3.1.3.11)
 DE (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPase).
 GN FBP.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Pisum.
 OX NCBI_TaxID=3888;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. GIANT; TISSUE=leaf;
 RX MEDLINE=95175626; PubMed=7870839;
 RA Dong S.M., Rhim J.H., Hahn T.R.;
 RT "Nucleotide sequence analysis of a cDNA encoding chloroplastic fructose-1,6-bisphosphatase from pea (Pisum sativum L.).";
 RL Plant Physiol. 107:313-314(1995).
 RN [2]
 RP SEQUENCE OF 27-407 FROM N.A.
 RC STRAIN=CV. LINCOLN; TISSUE=leaf;
 RX MEDLINE=94297517; PubMed=7764999;
 RA Carrasco J.L., Chuca A., Prado F.E., Hermoso R., Lazaro J.J.,
 RA Ramos J.L., Sahrawy M., Lopez Gorge J.;
 RT "Cloning, structure and expression of a pea cDNA clone coding for a photosynthetic fructose-1,6-bisphosphatase with some features different from those of the leaf chloroplast enzyme.";
 RT Planta 193:494-501(1994).
 RL [2]
 CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 1,6-BISPHOSPHATE + H(2O) -> D-FRUCTOSE 6-PHOSPHATE + ORTHOPHOSPHATE.
 CC -1- PATHWAY: THE CHLOROPLAST ISOZYME TAKES PART IN THE REGENERATION OF RIBULOSE BISPHOSPHATE IN THE PHOTOSYNTHETIC CARBON REDUCTION CYCLE (CALVIN CYCLE).
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
 CC -1- INDUCTION: LIGHT ACTIVATION THROUGH PH CHANGES, MG(2+) LEVELS AND ALSO BY LIGHT-MODULATED REDUCTION OF ESSENTIAL DISULFIDE GROUPS VIA THE FERREDOXIN-THIOREDOXIN F SYSTEM (BY SIMILARITY).
 CC -1- MISCELLANEOUS: IN PLANTS THERE ARE TWO FBPASE ISOZYMES: ONE IN THE CYTOSOL AND THE OTHER IN THE CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE FBPASE FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE FBPASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: U34806; AAD10213.1; -
 DR EMBL: X68826; CAA48719.1; -
 DR HSP; P22418; 1SP1.

```

DR InterPro: IPR000146; -
DR Pfam; PF00316; FBPHASE; 1.
DR PRINTS; PR00115; FBPHPTASE.
DR PRINTS; PR00377; INBPHPTASE.
DR PROSITE; PS00124; FBPHASE; 1.
KM Hydrolyase; Carbohydrate metabolism; Multigene family; Chloroplast;
KW Transit peptide; Calvin cycle.
FT TRANSIT 1 50
FT CHAIN 51 407
FT ACT_SITE 349 349
FT DISULFID 223 228
FT FT
FT FT
FT CONFLICT 82 82
FT CONFLICT 160 160
FT CONFLICT 247 247
FT CONFLICT 282 282
SO SEQUENCE 407 AA; 44511 MW; B60E9164F1FEFEDD CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 407;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 SMLGTG 135
DB 112 SMLGTG 119

RESULT 4
OCRA_MYCTU STANDARD; PRT; 429 AA.
ID OCRA_MYCTU
AC Q10387;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT (RIESKE IRON-
DE SULFUR PROTEIN).
GN OCRA OR RV2195 OR MTCY190.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX. THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF THE UBIQUINOL-CYTOCHROME C COMPLEX
CC ARE: CYTOCHROME B, THE RIESKE PROTEIN AND A CYTOCHROME C SUBUNIT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
CC BACTERIAL, CHLOROPLAST).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; Z70283; CAA94264.1; -
DR TuberculoList; RV2195; -
DR InterPro; IPR001281; -
DR PROSITE; PS00199; RIESKE_1; FALSE_NEG.
DR PROSITE; PS00200; RIESKE_2; 1.
KW Electron transport; Iron-sulfur; Oxidoreductase; Transmembrane.
FT TRANSMEM 96 116
FT TRANSMEM 137 157
FT TRANSMEM 207 227
FT METAL 353 353
FT METAL 355 355
FT METAL 372 372
FT METAL 375 375
FT DISULFID 358 374
SO SEQUENCE 429 AA; 46922 MW; C22C87EC283FE0DE CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 429;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 RKGOESFN 237
DB 335 RKGOESFN 342

RESULT 5
HPN_HELPY STANDARD; PRT; 59 AA.
ID HPN_HELPY
AC Q48251;
DT 15-JUL-1997 (Rel. 35, Created)
DT 15-JUL-1997 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HISTIDINE-RICH, METAL BINDING POLYPEPTIDE.
GN HPN OR HP1427 OR JHP1320.
OS Helicobacter pylori (Campylobacter pylori), and
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210, 85963;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18 AND 47-59.
RC STRAIN=LEU;
RX MEDLINE=95310028; PubMed=7790085;
RA Gilbert J.V., Ramakrishna J., Sunderman F.W. Jr., Wright A.,
RA Plant A.G.;
RT "Protein Hpn: cloning and characterization of a histidine-rich metal-
RT binding polypeptide in Helicobacter pylori and Helicobacter
RT mustelae."
RL Infect. Immun. 63:2682-2688(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McEnney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Bordovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 386:539-547(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deGorge B.L., Carnel G.,
RA Tumbino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,

```

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen *Helicobacter pylori*.";
 RL Nature 397:176-180(1999).
 CC -1- FUNCTION: STRONGLY BINDS NICKEL AND ZINC. BINDS OTHER METALS LESS
 CC STRONGLY: CO2+ > CU2+ > CD2+ > MN2+. MAY ACT TO INCREASE, OR AT
 CC LEAST TO PRESERVE, UREASE ACTIVITY. EXACT FUNCTION IS STILL
 CC UNKNOWN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U26361; AA05859.1; -
 CC EMBL: AE000643; AAD08471.1; -
 CC EMBL: AE001555; AAD06898.1; -
 CC TIGR: HP1427; -
 CC KM Metal-binding: Zinc; Nickel; Repeat.
 CC INIT_MER 0 0
 CC FT DOMAIN 10 23 POLY-HIS.
 CC FT DOMAIN 27 32 POLY-HIS.
 CC FT DOMAIN 37 54 2 X 5 AA REPEATS OF E-E-G-C-C.
 CC FT REPEAT 37 41 1.
 CC FT REPEAT 50 54 2.
 CC SO SEQUENCE 59 AA; 6946 MW; C3AE3F602EC973C CRC64;
 CC -----
 CC Query Match 2.6%; Score 7; DB 1; Length 59;
 CC Best Local Similarity 100.0%; Pred. No. 4.5;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 261 EHHNNH 267
 CC DB 26 EHHNNH 32
 CC -----
 CC RESULT 6
 CC ID YWKF_BACSU STANDARD; PRT; 95 AA.
 CC AC P45874;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE HYPOTHETICAL 10.1 KDA PROTEIN IN PFA-SPOIR INTERGENIC REGION.
 CC YWKF.
 CC OS *Bacillus subtilis*.
 CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC OC *Bacillus/staphylococcus* group; *Bacillus*.
 CC OX NCBI_Taxid=1423;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=168;
 CC RA Glaser P., Danchin A.;
 CC RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 CC RN [2]
 CC RP SEQUENCE OF 75-95 FROM N.A.
 CC RC STRAIN=168;
 CC RA Glaser P., Danchin A.;
 CC RL Submitted (OCT-1994) to the EMBL/Genbank/DBJ databases.
 CC CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC -----
 CC DR EMBL: Z49782; CAAB8986.1; -
 CC DR EMBL: Z38002; -; NOT ANNOTATED_CDS.
 CC DR EMBL: Z99122; CAB15716.1; -
 CC DR Subtilisin; BG11317; ywKF.
 CC KW Hypothetical protein; Transmembrane.
 CC FT TRANSMEM 5 25 POTENTIAL.
 CC FT TRANSMEM 31 51 POTENTIAL.
 CC FT TRANSMEM 75 95 POTENTIAL.
 CC SO SEQUENCE 95 AA; 10069 MW; 2A2F80BDEC705BC4 CRC64;
 CC -----
 CC Query Match 2.6%; Score 7; DB 1; Length 95;
 CC Best Local Similarity 100.0%; Pred. No. 6.8;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 26 RMKAAA 32
 CC DB 71 RMKAAA 77
 CC -----
 CC RESULT 7
 CC ID YGH2_YEAST STANDARD; PRT; 119 AA.
 CC AC P53161;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
 CC DE HYPOTHETICAL 13.4 KDA PROTEIN IN HSF1-AFT1 INTERGENIC REGION.
 CC YGH2_YEAST.
 CC OS *Saccharomyces cerevisiae* (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CC OX NCBI_Taxid=4932;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=S288C;
 CC RX MEDLINE=97435481; PubMed=9290212;
 CC RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
 CC RT "Sequence analysis of 203 kilobases from *Saccharomyces cerevisiae*
 CC RT chromosome VII.";
 CC RL Yeast 13:1077-1090(1997).
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: Z72596; CAA96778.1; -
 CC DR SGD: S0003040; YGH072C.
 CC KW Hypothetical protein; Transmembrane.
 CC FT TRANSMEM 30 50 POTENTIAL.
 CC SO SEQUENCE 119 AA; 13388 MW; F1D2F040C1C833A6 CRC64;
 CC -----
 CC Query Match 2.6%; Score 7; DB 1; Length 119;
 CC Best Local Similarity 100.0%; Pred. No. 8.3;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 7 FSGAELY 13
 CC DB 98 FSGAELY 104
 CC -----
 CC RESULT 8
 CC ID NODP_RHIS3 STANDARD; PRT; 301 AA.
 CC AC P72338;
 CC DT 30-MAY-2000 (Rel. 39, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SULFATE ADENYLYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLYLTRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT) (MODULATION PROTEIN P.).
 DE P.).
 GN NODP.
 OS Rhizobium sp. (strain N33).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=103798;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cloutier J.;
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER TO NOD FACTOR.
 CC -1- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE + ADENYLYLSULFATE.
 CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYMES, NODP AND NODQ, MAY BE PHYSICALLY ASSOCIATED (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY, CYSD SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U53327; AAB16901.1; -;
 DR InterPro; IPR002500; -;
 DR Pfam; PF01507; PAPS_reduct; 1.
 KW Modulation; Transferase; Nucleotidyltransferase.
 SO SEQUENCE 301 AA; 35008 MW; 3582F7EAB0861431 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 301;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 220 RDGMIL 226
 |||||
 DB 218 RDGMIL 224

RESULT 9
 HEMZ_BRAJA STANDARD; PRT; 345 AA.
 AC P28602;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE FERROCHELATASE (EC 4.99.1.1) (PROTOHEME FERRO-LYASE) (HEME SYNTHETASE).
 GN HEMH.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LO;
 RX MEDLINE=92325004; PubMed=1624416;
 RA Frustaci J.M., O'Brian M.R.;
 RT "Characterization of a Bradyrhizobium japonicum ferrochelatase mutant and isolation of the hemh gene";
 RL J. Bacteriol. 174:4223-4229(1992).
 RN [2]
 RP REVISIONS.
 RA O'Brian M.;
 RL Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX. ESSENTIAL FOR NORMAL NODULE DEVELOPMENT.

CC -1- CATALYTIC ACTIVITY: PROTOPORPHYRIN + FE(2+) -> PROTOHEME + 2 H(+).
 CC -1- PATHWAY: LAST STEP IN PROTOHEME BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FERROCHELATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M92427; AAA6217.2; -;
 DR PIR; A42883; A42883.
 DR InterPro; IPR001015; -;
 DR Pfam; PF00762; Ferrochelatase; 1.
 DR PROSITE; PS00534; FERROCHELATASE; 1.
 KW Porphyrin biosynthesis; Heme biosynthesis; Lyase; Iron.
 FT METAL 215 215
 FT METAL 296 296
 SO SEQUENCE 345 AA; 38431 MW; 23E1658F40D38788 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 345;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 254 DKLAAL 260
 |||||
 DB 102 DKLAAL 108

RESULT 10
 HRPX_PLALO STANDARD; PRT; 351 AA.
 ID HRPX_PLALO
 AC P04929;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1989 (Rel. 38, Last annotation update)
 DE HISTIDINE-RICH GLYCOPROTEIN PRECURSOR.
 OS Plasmodium lophurae.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5853;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85061618; PubMed=6095114;
 RA Ravetch J.V., Feder R., Pavlovic A., Blobel G.;
 RT "Primary structure and genomic organization of the histidine-rich protein of the malaria parasite Plasmodium lophurae";
 RL Nature 312:616-620(1984).
 CC -1- MISCELLANEOUS: IN THE INTRABERYTHROCYTIC STAGES OF DEVELOPMENT OF P. LOPHURAE IN DUCKS, THERE IS A SYNTHESIS OF A MAJOR PROTEIN THAT ACCUMULATES TO COMPRISE AT LEAST 50% OF THE CELLULAR MASS: THE HISTIDINE RICH PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X01469; CAA25698.1; -;
 DR PIR; A22692; KGZOH.
 KW Malaria; Repeat; Glycoprotein; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 47
 FT CHAIN 48 351
 FT CARBOHYD 40 40
 FT DOMAIN 59 90
 FT REPEAT 59 74
 FT HISTIDINE-RICH GLYCOPROTEIN
 N-LINKED (GLCNAC...) (PROBABLE).
 2 X 16 AA TANDEM REPEATS.
 16-1.

FT REPEAT 75 90 16-2.
 FT DOMAIN 91 123 2 x 17 AA TANDEM REPEATS.
 FT REPEAT 91 107 17-1.
 FT REPEAT 108 123 17-2.
 FT DOMAIN 124 153 2 x 15 AA TANDEM REPEATS.
 FT REPEAT 124 138 15-1.
 FT REPEAT 139 153 15-2.
 FT DOMAIN 173 351 18 x 10 AA TANDEM REPEATS.
 SQ SEQUENCE 351 AA; 44032 MW; D19A48D47D890453 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 351;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHHHH 267
 |||||
 Db 132 EHHHHH 138

RESULT 11
 MIX2_XENLA STANDARD; PRT; 369 AA.
 ID 091685;
 AC 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HOMEOBOX PROTEIN MIX.2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 NCBI_TaxID=8355;
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=9629435; Pubmed=8660890;
 RA Vize P.D.;
 RT "DNA sequences mediating the transcriptional response of the Mix.2
 RT homeobox gene to mesoderm induction."
 RL Dev. Biol. 177:226-231(1996).
 CC -1- FUNCTION: MAY BE A TRANSCRIPTION FACTOR WHICH PLAY A REGULATORY
 CC ROLE IN THE DEVELOPMENT OF THE EMBRYO (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- INDUCTION: BY ACTIVIN AND TGF-BETA1 (IMMEDIATE EARLY RESPONSE
 CC GENE).
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
 CC
 CC EMBL: 050745; AAC60020.1; -
 CC DR HSSP: P06601; 1FJL.
 CC DR InterPro: IPR001356; -
 CC DR Pfam: PF00046; homeobox.1.
 CC DR PROSITE: PS00027; HOMEOBOX_1; 1.
 CC DR PROSITE: PS0071; HOMEOBOX_2; 1.
 CC KW Homeobox: DNA-binding; Developmental protein; Nuclear protein.
 CC FT DNA_BIND 94 153 HOMEOBOX.
 CC SQ SEQUENCE 369 AA; 41537 MW; FC02182C85D6B9F5 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 369;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 246 PAPNSSS 252
 |||||
 Db 178 PAPNSSS 184

RESULT 12
 MIX1_XENLA STANDARD; PRT; 377 AA.
 AC P21711;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HOMEOBOX PROTEIN MIX.1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 NCBI_TaxID=8355;
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=89288302; Pubmed=2567635;
 RA Rosa F.M.;
 RT "Mix.1, a homeobox mRNA inducible by mesoderm inducers, is expressed
 RT mostly in the presumptive endodermal cells of Xenopus embryos."
 RL Cell 57:965-974(1989).

CC -1- FUNCTION: MAY BE A TRANSCRIPTION FACTOR WHICH PLAY A REGULATORY
 CC ROLE IN THE DEVELOPMENT OF THE EMBRYO. INVOLVED IN THE
 CC ESTABLISHMENT OF DORSAL/VENTRAL PATTERN IN THE EARLY MESODERM.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MOSTLY IN THE PRESUMPTIVE ENDODERMAL
 CC CELLS OF XENOPUS EMBRYOS.
 CC -1- INDUCTION: BY ACTIVIN AND TGF-BETA1 (IMMEDIATE EARLY RESPONSE
 CC GENE).
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sid.ch).
 CC
 CC EMBL: M21063; AAA49903.1; -
 CC DR PIR: A32548; A32548.
 CC DR HSSP: P06601; 1FJL.
 CC DR InterPro: IPR001356; -
 CC DR Pfam: PF00046; homeobox.1.
 CC DR PROSITE: PS00027; HOMEOBOX_1; 1.
 CC DR PROSITE: PS0071; HOMEOBOX_2; 1.
 CC KW Homeobox: DNA-binding; Developmental protein; Nuclear protein.
 CC FT DNA_BIND 96 155 HOMEOBOX.
 CC FT DOMAIN 352 368 ASP/GLU-RICH (ACIDIC).
 CC SQ SEQUENCE 377 AA; 42250 MW; 6BA404359B57D221 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 377;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 246 PAPNSSS 252
 |||||
 Db 183 PAPNSSS 189

RESULT 13
 HMPA_ECOLI STANDARD; PRT; 396 AA.
 ID P24232;
 AC P24232;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE FLAVOHEMOPROTEIN (HEMOGLOBIN-LIKE PROTEIN) (FLAVOHEMOGLOBIN)
 DE (DIHYDROPTERIDINE REDUCTASE (EC 1.6.99.7)) (FERRISIDEROPHORE REDUCTASE
 DE B) (NITRIC OXIDE DIOXYGENASE) (NOD).
 GN HMP OR HMPA OR FSRB.

OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_Taxid=562;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN-K12;
RX MEDLINE=91238719; PubMed=2034230;
RA Vasudevan S.G., Armarego W.L.F., Shaw D.C., Lilley P.E., Dixon N.E.,
RA Poole R.K.;
RT "Isolation and nucleotide sequence of the hmp gene that encodes a
RT haemoglobin-like protein in Escherichia coli K-12.";
RL Mol. Gen. Genet. 226:49-58(1991).
RN (12)
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97456617; PubMed=9278503;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Osima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT DNA Res. 4:91-113(1997).
RN (14)
RP SEQUENCE OF 1-10 FROM N.A.
RX MEDLINE=83235562; PubMed=6190704;
RA Pliamun M.D., Stauffer G.V.;
RT "Characterization of the Escherichia coli gene for serine
RT hydroxymethyltransferase.";
RL Gene 22:9-18(1983).
RN (15)
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC STRAIN-K12;
RX MEDLINE=92290008; PubMed=1601132;
RA Andrews S.C., Shipley D., Keen J.N., Findlay J.B.C., Harrison P.M.,
RA Guest J.R.;
RT "The haemoglobin-like protein (HMP) of Escherichia coli has
RT ferrisulphate reductase activity and its C-terminal domain shares
RT homology with ferredoxin NADP+ reductases.";
RL FEBS Lett. 302:247-252(1992).
RN (16)
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EWG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1311(1997).
RN (17)
RP SEQUENCE OF 1-20, AND CHARACTERIZATION.
RC MEDLINE=9833652; PubMed=9724711;
RA Gardner P.R., Gardner A.M., Martin L.A., Salzman A.L.;
RT "Nitric oxide dioxygenase: an enzymic function for flavohemoglobin.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10378-10383(1998).
CC -1- FUNCTION: POSSESSES DIHYDROPTERIDINE REDUCTASE ACTIVITY. ALSO HAS
CC NITRIC OXIDE DIOXYGENASE ACTIVITY.
CC -1- FUNCTION: SEEMS TO REDUCES THE COMPLEXED FE+3 IRON OF SIDEROPHORES
CC TO FE+2. THUS RELEASING IT FROM THE CHELATOR.
CC -1- CATALYTIC ACTIVITY: NAD(P)H + 6,7-DIHYDROPTERIDINE = NAD(P)(+) +

```
CC      5, 6,7,8-TETRAHYDROPTERIDINE.
CC      -1- SUBUNIT: MONOMER.
CC      -1- DOMAIN: CONSISTS OF TWO DISTINCT DOMAINS; ONE IS A HEME-CONTAINING
CC      OXYGEN BINDING DOMAIN IN THE N-TERMINAL REGION AND THE OTHER IS AN
CC      FAD-CONTAINING REDUCTASE DOMAIN FOUND IN THE C-TERMINAL REGION..
CC      -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY. TWO DOMAINS
CC      FLAVOHEMOPROTEINS SUBFAMILY.
CC      -1- SIMILARITY: TO A NUMBER OF FAD/NAD(P) FLAVOPROTEIN
CC      OXIDOREDUCTASES.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X58872; CAA41682.1; -.
DR      EMBL: AE000341; AAC75605.1; -.
DR      EMBL: D90885; BAU16460.1; -.
DR      EMBL: J01620; AAA23911.1; -.
DR      PIR: S15992; S15992.
DR      PIR: S21161; S21161.
DR      HSSP: P04252; IVHB.
DR      Ecogene; EG10456; hmp.
DR      InterPro; IPR000971; -.
DR      InterPro; IPR001433; -.
DR      Pfam; PF00042; globin; 1.
DR      Pfam; PF00175; oxidored_fad; 1.
DR      PROSITE; PS01033; GLOBIN; 1.
KW      Oxidoreductase; NADP; Heme; Flavoprotein; FAD; Iron transport.
FT      METAL
FT      METAL          53      134
FT      METAL          53      53
FT      METAL          85      85
FT      NP_BIND        268     273
FT      NP_BIND        396 AA; 43867 MW; 49961BDE1444BD68 CRC64;
SO      SEQUENCE

Query Match           2.6%; Score 7; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. NO. 24;
Matches    7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      123 EGGGVSN 129
       |||||
Db      225 EGGGVSN 231

RESULT 14
YDJN_HAEIN
AC      PA5079; STANDARD; PRT; 440 AA.
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      HYPOTHEITICAL SYMPORFER H1154.
GN      H1154.
OS      Haemophilus influenzae.
OC      Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC      Haemophilus.
OX      NCBI_TaxID=727;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-RD / KW20 / ATCC 51907.
RX      MEDLINE=95350630; Pubmed=7542800;
RA      Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA      Kelleys A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA      McEneney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA      Scott J.D., Shiley R., Liu L.-I., Glock A., Kelley J.M.,
RA      Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA      Uetendack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
```

RA Fine L.D., Fritschman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.:
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT Influenzae Rd".
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE SODIUM-DICARBOXYLATE SYMPORTEF FAMILY
 CC (SDF). STRONG, TO E.COLI YDUN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32795; AAC22809.1; -.
 DR TIGR: H1154; -.
 DR InterPro: IPR001991; -.
 DR Pfam: PF00375; SDF: 1.
 DR PROSITE: PS00713; NA-DICARBOXYL_SYM_1; FALSE_NEG.
 DR PROSITE: PS00714; NA-DICARBOXYL_SYM_2; FALSE_NEG.
 KW Hypothetical protein; Transmembrane; Transport.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 29 49 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.
 FT TRANSMEM 343 363 POTENTIAL.
 FT TRANSMEM 366 386 POTENTIAL.
 FT TRANSMEM 389 409 POTENTIAL.
 SQ SEQUENCE 440 AA; 46491 MW; F929839559FD06A CRC64;

Query Match 2.6%; Score 7; DB 1; Length 440;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GOVSNLT 131
 IIIIIIII
 DB 150 GOVSNLT 156

RESULT 15
 ETS2_MOUSE
 ID ETS2_MOUSE STANDARD; PRT; 468 AA.
 AC P15037;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE C-ETS-2 PROTEIN.
 GN ETS2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89042086; PubMed=2847145;
 RA Watson D.K., McWilliams M.J., Lapsis P., Lautenberger J.A.,
 RA Schweinfest C.W., Papas T.S.;
 RT "Mammalian ets-1 and ets-2 genes encode highly conserved proteins".
 RC Proc. Natl. Acad. Sci. U.S.A. 85:7862-7866(1988).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J04103; AAA37581.1; -.
 DR PIR: C32066; TWSE2.
 DR HSSP: P14921; 2STW.
 DR TRANSFAC: T01397; -.
 DR MGD: MGI:95456; Ets2.
 DR InterPro: IPR000418; -.
 DR Pfam: PF00178; Ets; 1.
 DR PRINTS: PR00454; ETSDOMAIN.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS0061; ETS_DOMAIN_3; 1.
 KW Proto-oncogene; DNA-binding; Nuclear protein.
 FT DOMAIN 87 170 POINTED.
 FT DNA_BIND 362 442 ETS-DOMAIN.
 SQ SEQUENCE 468 AA; 52827 MW; 5260F3085B7EB831 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 468;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 190 FDFPDGS 196
 IIIIIIII
 DB 29 FDFPDGS 35

Search completed: August 23, 2001, 11:18:28
 Job time: 195 sec


```

Db      77 tspsynkfdcgensvqlygylyevrtnkpkaknlgivssftlycpte---gtrpwdeidi 133
QY      83 EVLGKKNPGSFGSNIIITGKAGAKQKTSSEKHHA VSPADDAQFHTYGLFEMPTNYRVKRWYVDGGEV 142
      134 eflgkdtlcvfnytlngvgqh---ekvislqfaskgfhlyafdwgpylykwyvdg-vl 189
QY      143 RKTGGGVSNLTGTQG-LRFNLMSSESA-AWVGOPDESKLPFQFINWVKYKKT 195
      190 kht---atanlpsrpgkimmlwngtgyvdwlgysyngan-plyaeaydwv---kyl 237

```

RESULT 4

```

AAR03775 standard; protein: 234 AA.
ID      AAR03775
XX
AC      AAR03775;
XX
DT      31-JUL-1990 (first entry)
XX
DE      Thermostable beta-glucanase.
XX
KM      Thermostable beta-glucanase; ss; Bacillus subtilis.
XX
OS      Bacillus macerans.
XX
PN      DD272102-A.
XX
PD      27-SEP-1989.
XX
PF      12-MAY-1988; 88DD-0315706.
XX
PR      12-MAY-1988; 88DD-0315706.
XX
PA      (DEAK ) AKAD WISSENSCHAFT DDR.
XX
PI      Borriess R;
XX
DR      WPI; 1990-067913/10.
XX
N-PSDB; AAQ03519.
XX
PT      Thermostable beta-glucanase production -
PT      using Bacillus subtilis transformed with gene from Bacillus
PT      macerans.
XX
PS      Fig 1. ; 9pp: German.
XX
CC      The gene encoding Bacillus macerans thermostable
CC      beta-glucanase is expressed in Bacillus subtilis. The
CC      enzyme is useful for lowering the viscosity of brewing
CC      mashes and in the production of feedstuff.
XX
SQ      Sequence 234 AA;

```

Query Match

Best Local Similarity 12.5%; Score 228; DB 11; Length 234;

Matches 60; Conservative 25; Mismatches 74; Indels 16; Gaps 8;

```

QY      23 TNVSAKDFSGAELTYLLEVOYGFKEARMKMAAAGTVSSMFLYONGSEIADGRPWVEVDI 82
      72 tssaynkfdcaeyrstnlygylvevsmkpkaknlgivssftlycpr---ahgtqwdelid 128
Db      83 EVLGKKNPGSFGSNIIITGKAGAKQKTSSEKHHA VSPADDAQFHTYGLFEMPTNYRVKRWYVDGGEV 142
      129 eflgkdtlcvfnytlngvgqh---ekvislqfaskgfhlyafdwgpylykwyvdg-vl 184
QY      143 RKTGGGVSNLTGTQG-LRFNLMSSESA-AWVGOPDESKLPFQFINWVKYKKT 195
      185 kht---atanlpsrpgkimmlwngtgyvdwlgysyngan-plyaeaydwv---kyl 232
Db

```

RESULT 5

```

AAR05803
ID      AAR05803 standard; protein: 237 AA.
XX
AC      AAR05803;
XX
DT      08-NOV-1990 (first entry)
XX
DE      Heat-stable endo-beta-1,3-1,4-glucanase.
XX
KM      Endo-beta-1,3-1,4-glucanase; barley; brewing.
XX
OS      Bacillus macerans.
XX
PN      DD275704-A.
XX
PD      31-JAN-1990.
XX
PF      23-SEP-1988; 88DD-0320082.
XX
PR      23-SEP-1988; 88DD-0320082.
XX
PA      (DEAK ) AKAD WISSENSCHAFT DDR.
XX
PI      Borriess R, Wobus U, Mendel R-R, Baumlein H;
XX
DR      WPI; 1990-210631/28.
XX
N-PSDB; AAQ05167.
XX
PT      Prep'n. of barley plants expressing heat stable beta-glucanase -
PT      by transforming cells with appropriate vector then regeneration
PT      giving seeds useful in brewing without conversion to malt.
XX
PS      Disclosure; ; p; German.
XX
CC      The corresp. gene and the plant promoter region are inserted into an
CC      expression vector and used to transform barley cells. The transformants
CC      are used to regenerate barley plants are useful in brewing. This
CC      protein accumulates in the ripe (but ungerminated) seeds. The
CC      proportion of these to malted seeds is increased therefore, without
CC      an unacceptable increase in viscosity.
XX
SQ      Sequence 237 AA;

```

Query Match

Best Local Similarity 12.5%; Score 228; DB 11; Length 237;

Matches 60; Conservative 25; Mismatches 74; Indels 16; Gaps 8;

```

QY      23 TNVSAKDFSGAELTYLLEVOYGFKEARMKMAAAGTVSSMFLYONGSEIADGRPWVEVDI 82
      75 tssaynkfdcaeyrstnlygylvevsmkpkaknlgivssftlycpr---ahgtqwdelid 131
Db      83 EVLGKKNPGSFGSNIIITGKAGAKQKTSSEKHHA VSPADDAQFHTYGLFEMPTNYRVKRWYVDGGEV 142
      132 eflgkdtlcvfnytlngvgqh---ekvislqfaskgfhlyafdwgpylykwyvdg-vl 187
QY      143 RKTGGGVSNLTGTQG-LRFNLMSSESA-AWVGOPDESKLPFQFINWVKYKKT 195
      188 kht---atanlpsrpgkimmlwngtgyvdwlgysyngan-plyaeaydwv---kyl 235
Db

```

RESULT 6

AAR06622 standard; protein: 237 AA.

ID AAR06622;

AC AAR06622;

DT 09-JAN-1991 (first entry)

DE Hybrid (1,3-1,4)-pre-beta-glucanase.

XX Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry.

OS Bacillus amyloliquefaciens, Bacillus macerans.
 XX Key Location/Qualifiers
 FH Domain 1..127
 FT Domain /label=amln- terminal half of B.macerans beta-glucanase
 FT Domain 131..234
 FT Domain /label=carboxyl- terminal half of B.amyloliquefaciens
 FT Domain beta-glucanase
 PN MO9009436-A.
 XX
 PD 23-AUG-1990.
 XX
 PF 16-FEB-1990; 90MO-DK00044.
 XX
 PR 04-AUG-1989; 89DK-0003848.
 XX 16-FEB-1989; 89DD-0325800.
 XX
 PA (CARL-) CARLSBERG A/S.
 PA (DEAK-) AKAD WISSENSCHAFT DDR.
 PI Boriss R, Hofemeister J, Thomsen KK, Olsen O, Vonwetstein D;
 XX MPI; 1990-275129/36.
 DR N-PSDB; AAO05833.
 XX
 PT New thermostable (1,3-1,4)-beta-glucanase - prep'd using hybrid
 PT gene obtd. using Bacillus amyloliquefaciens and B.macerans genes
 XX
 PS Disclosure: page 28; 84pp; English.
 XX
 CC This hybrid protein is encoded by the beta-glucanase-H2 gene.
 CC Following processing of the signal peptide the mature protein
 CC is produced, comprising the amino terminus of the B.macerans
 CC beta-glucanase and the carboxyl-terminal half of the B.amylo-
 CC liquefaciens beta-glucanase. This hybrid protein is thermo-
 CC stable and hydrolyses beta-glycosidic linkages in (1,3-1,4)-
 CC beta-glucans. Reducing sugars are obtd. at high temps. and
 CC thus this enzyme can be used in the mfr. of food prods., esp.
 CC beer and animal feed (e.g. for feeding poultry). See also
 CC AAO05833.
 CC
 SQ Sequence 237 AA;

Query Match 12.2%; Score 222.5; DB 11; Length 237;
 Best Local Similarity 33.1%; Pred. No. 5.5e-11;
 Matches 57; Conservative 25; Mismatches 79; Indels 11; Gaps 6;
 OY 23 TNNVSAKDFSGAELTYLLEVOYGFKFEARKMAAASGTVSSMFLYONGSEIADGRPWEVDI 82
 DB 75 tsaaykctfdcaeyrtnlygylyevsmkpkatgtlvssfftytgp---ahgfwelidl 131
 OY 83 EVLGNKPGSQSNITITGKAGAKQKTSKHNHVASPAADQAFHTYGLWTPNRYVKTVDGQEV 142
 DB 132 eflgkctctkvfnyltnngag---nhkfadlqfdaanaylytafdqgpnstikvyvdgq-1 187
 OY 143 RRTGGOVSNLTGCTGRLFLMSSESA-AMVGOFDESKLPLFOFINMVKYK 193
 DB 188 khtatqdlpaapqg---immlmwtgtdvddwlgsgnvgv-plyahymwtyrk 236

RESULT 7
 AAM37884
 ID AAM37884 standard; Protein; 245 AA.
 AC AAM37884;
 XX
 DT 20-AUG-1998 (first entry)
 XX
 DE Lichenase protein.
 XX
 KW Lichenase; lica; fungus; enzyme; beta-1,4-glucan bond hydrolysatlon;

KM beta-1,3-linked glucan bond; grain-containing feed; grain treatment;
 KM nutrient availability; brewing; fermentation.
 XX
 OS Orpinomyces sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..29
 FT Peptide /note="signal peptide"
 FT Protein 30..246
 FT Protein /note="mature lichenase"
 PN MO9814595-A1.
 XX
 PD 09-APR-1998.
 XX
 PF 03-OCT-1997; 97MO-US17811.
 XX
 PR 04-OCT-1996; 96US-0027882.
 XX
 PA (UYGP-) UNIV GEORGIA RES FOUND INC.
 PA Chen H, Li X, Ljungdahl LG;
 PI
 XX MPI; 1998-240094/21.
 DR N-PSDB; AAV29067.
 XX
 PT New isolated lichenase protein - is obtained from Orpinomyces PC-2,
 PT used for treatment of grain to improve feeds or to improve brewing
 and fermentation processes
 XX
 PS Claim 1; Page 24-25; 41pp; English.
 XX
 CC This sequence is the lichenase (lica) of Orpinomyces sp. strain PC-2
 CC of the invention. The protein was purified from a fungus or a fungal
 CC culture or from a recombinant DNA molecule having a fungal lichenase
 CC coding sequence. The lichenase enzyme hydrolyses the beta-1,4-glucan
 CC bonds adjacent to the beta-1,3-linked glucan bonds, but does not cleave
 CC beta-1,4-linked glucans. They can be used for the treatment of animal
 CC grain-containing feeds to improve nutrient availability and for treatment
 CC of grain (e.g. barley or wheat) in the brewing and fermentation
 CC industries to increase carbon substrate availability and to maximise
 CC production of desired products.
 CC
 SQ Sequence 245 AA;

Query Match 10.4%; Score 190.5; DB 19; Length 245;
 Best Local Similarity 31.6%; Pred. No. 2.8e-08;
 Matches 54; Conservative 17; Mismatches 87; Indels 13; Gaps 6;
 OY 21 LFTNVSAKDFSGAELTYLLEVOYGFKFEARKMAAASGTVSSMFLYONGSEIADGRPWEV 80
 DB 79 lldrdgsgyltgeyrtknyygmfgvnmkpklnpnyvssfftytgps---dgtkwdei 135
 OY 81 DIEVLAKNPGSQSNITITGKAGAKQKTSKHNHVASPAADQAFHTYGLWTPNRYVKTVDGQ 140
 DB 136 dleflgdtctkvfnyltnngqgh---enhlylgfdaagfbygtfawrnstlwyvdc 192
 OY 141 EVKRTGGOVSNLTGCTG-RLFLMSSESA-AMVGOFDESKLPLFOFINMV 189
 DB 193 ayla-----ydnldtprklnmawnglgvddwlrpfm-grtnlsayvdcw 238

RESULT 8
 AAG48467
 ID AAG48467 standard; Protein; 282 AA.
 AC AAG48467;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 61207.
 XX

Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EPI033405-A2.
XX
XX
PD 06-SEP-2000.
PD
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134370.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135121.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139753.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.

PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.

PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155653.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0158293.
PR	13-OCT-1999;	99US-0158294.
PR	13-OCT-1999;	99US-0158295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	10.1%	Score 185	DB 21	Length 282
Best Local Similarity	27.4%	Pred. No. 9	9e-08	
Matches 59	Conservative	17	Mismatches 96	Indels 24
Gaps				
QY	24	NVSAKDFSGAEILYLTLEEVOYGFKAEMKMAA--ASGTVSSMFLYONGSEIADGRPWEVD	81	
Db	53	slslldkftsgsgfsgshqeflygkvevqmkllvpngasgvtlyflk-----spgtlweid	106	
QY	82	IEVLGK---NGSGFQSNLITGKAGAQKISEKHNAVSRAADQAFHNYGLEMTNYYRWYTD	138	
Db	107	feflgnlsghpyclhtnvyt-kgvtdkdaqfhlwfdrrpvd-fhygclimrpqvlvftld	163	
QY	139	GOEVRKTEGGGVSNL--TGTOGLRF--NLMSSESAAWG---QFDESKLRLPQFINWAV	191	
Db	164	giprlrefnnsaelgyprfkhqgmrllyaslweehewatctgglektctwskarfcalfynynlv	223	
QY	192	YKTPGQSGEGSGDFTLMDTNDNEFTDGS---RWGK	223	
Db	224	dacvwsngksscasanswftgylvldfkxnyvkwaq	258	
RESULT	9			
AAG32464				
ID	AAG32464	standard; Protein: 282 AA.		
XX				
GC	AAG32464;			

XX	DT	17-OCT-2000	(first entry)
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 39167.	
XX	KW	Protein identification: signal transduction pathway; metabolic pathway;	
XX	KW	hybridisation assay; genetic mapping; gene expression control; promoter	
XX	KW	termination sequence.	
XX	OS	Arabidopsis thaliana.	
XX	PN	EP1033405-A2.	
XX	ED	06-SEP-2000.	
XX	PF	25-FEB-2000; 2000EP-0301439.	
XX	PR	25-FEB-1999; 99US-0121825.	
XX	PR	05-MAR-1999; 99US-0123180.	
XX	PR	09-MAR-1999; 99US-0123548.	
XX	PR	23-MAR-1999; 99US-0125788.	
XX	PR	25-MAR-1999; 99US-0126264.	
XX	PR	29-MAR-1999; 99US-0126785.	
XX	PR	01-APR-1999; 99US-0127452.	
XX	PR	06-APR-1999; 99US-0128234.	
XX	PR	08-APR-1999; 99US-0128714.	
XX	PR	16-APR-1999; 99US-0129845.	
XX	PR	19-APR-1999; 99US-0130077.	
XX	PR	21-APR-1999; 99US-0130449.	
XX	PR	23-APR-1999; 99US-0130510.	
XX	PR	23-APR-1999; 99US-0130891.	
XX	PR	28-APR-1999; 99US-0131449.	
XX	PR	30-APR-1999; 99US-0132048.	
XX	PR	04-MAY-1999; 99US-0132484.	
XX	PR	05-MAY-1999; 99US-0132485.	
XX	PR	06-MAY-1999; 99US-0132486.	
XX	PR	07-MAY-1999; 99US-0132487.	
XX	PR	11-MAY-1999; 99US-0134256.	
XX	PR	14-MAY-1999; 99US-0134218.	
XX	PR	14-MAY-1999; 99US-0134219.	
XX	PR	14-MAY-1999; 99US-0134221.	
XX	PR	14-MAY-1999; 99US-0134370.	
XX	PR	18-MAY-1999; 99US-0134768.	
XX	PR	19-MAY-1999; 99US-0134941.	
XX	PR	20-MAY-1999; 99US-0135124.	
XX	PR	21-MAY-1999; 99US-0135353.	
XX	PR	24-MAY-1999; 99US-0135629.	
XX	PR	25-MAY-1999; 99US-0136021.	
XX	PR	27-MAY-1999; 99US-0136392.	
XX	PR	28-MAY-1999; 99US-0136782.	
XX	PR	01-JUN-1999; 99US-0137222.	
XX	PR	03-JUN-1999; 99US-0137528.	
XX	PR	04-JUN-1999; 99US-0137502.	
XX	PR	07-JUN-1999; 99US-0137724.	
XX	PR	08-JUN-1999; 99US-0138040.	
XX	PR	10-JUN-1999; 99US-0138540.	
XX	PR	10-JUN-1999; 99US-0138847.	
XX	PR	14-JUN-1999; 99US-0139119.	
XX	PR	16-JUN-1999; 99US-0139452.	
XX	PR	16-JUN-1999; 99US-0139453.	
XX	PR	17-JUN-1999; 99US-0139492.	
XX	PR	18-JUN-1999; 99US-0139454.	
XX	PR	18-JUN-1999; 99US-0139455.	
XX	PR	18-JUN-1999; 99US-0139456.	
XX	PR	18-JUN-1999; 99US-0139457.	
XX	PR	18-JUN-1999; 99US-0139458.	
XX	PR	18-JUN-1999; 99US-0139459.	
XX	PR	18-JUN-1999; 99US-0139460.	
XX	PR	18-JUN-1999; 99US-0139461.	
XX	PR	18-JUN-1999; 99US-0139462.	
XX	PR	18-JUN-1999; 99US-0139463.	

Query Match	Best Local Similarity	Score	DB	Length
24 NVSAKDFGAEFLYEVQYGFARAKMA--ASGVSSMFLYONGSEIADQRPWEVD	81	9.58	173	282
53 sisldkssgsfgngelfygaevqmklypvnasatvttfyyk-----sptlwdeld	106	25.68	1e-06	12
Matches 64; Conservative 36; Mismatches 92; Indels 58; Gaps				
82 IEVLGK--NPGSFOSNIITGKAGAOCTSEKHNAVSPADDAFHLYGLEMTNRYKMTVD	138			
107 fefllngslgnphylthnlytc-ktglgdkcggfhlwfdpeyn--fhycclwmpqrlifftvd	163			
139 GQEVKRTIEGGGVSNL--TGTOGLRF--NLMSSSAAMVG---QFDSKRLPQFINWKV	191			
164 gipirefknpcaigvpifrlqpmrlyaslwaeahwalgldkxwskapfafe-----	217			
192 YKTRPGQEGSDFTLMTWNTDFPDGSRMGKGDWTFDGNRVDLFTKNTYSDGMILIAL	251			
218 YR-----hynv--dgcvwangksscsanspwftqk-----	246			
252 TRKGQESFNG	261			

Db 247 dsngqtrmkq 256

RESULT 10
AAG52114
ID AAG52114 standard; Protein: 277 AA.
XX
AC AAG52114;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66212.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; pldmoter;
KM termination sequence.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136592.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160881.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      9.5%; Score 172.5; DB 21; Length 277;
Best Local Similarity 26.6%; Pred. No. 1, le-06;
Matches 54; Conservative 34; Mismatches 74; Indels 41; Gaps 9;
```

```
OY 24 NVSADSGAELYLTLEVOYGFKEARMKAA--ASGVSSMFLYNGSEIADGRPWVEVD 81
Db 48 slsldkssgsgfsgnqeflylkaevqmklypnsagtytlylk-----spgltwdeid 101
OY 82 IEVLCK---NGSQFSNITITGKAGAKTSEKHAVSPAADAFHHTYGLFWPNNYVRWTVVD 138
Db 102 feflgnisghpylhtnvyt-kgsdqkqgfhlfwfdptan--fhcyeltwmpqrllftvd 158
```

```
OY 139 GQEVKRTREGQVSNLTGTQGLRF-----NLSSESAAWVG--QFDESKLPIRF 185
Db 159 gplrfefmae-----grgyfpkqpmrlyaslwseehwatrgjlektkswkaptfay 212
OY 186 IN-----WKVKYKTPGCGE 200
Db 213 yrnynvegcvwnvgksvcpsansq 235
```

RESULT 11

AA018644 ID AAG18644 standard; Protein; 269 AA.

AA018644; AC AAG18644;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 20135.

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

KM Arabidopsis thaliana.

OS Arabidopsis thaliana.

EP103405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130549.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134370.

PR 14-MAY-1999; 99US-0134371.

PR 18-MAY-1999; 99US-0134768.

PR 18-MAY-1999; 99US-0134941.

PR 19-MAY-1999; 99US-0135124.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137720.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142820.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.3%; Score 170.5; DB 21; Length 269;
Best Local Similarity 27.6%; Pred. No. 1.5e-06;
Matches 59; Conservative 35; Mismatches 97; Indels 23; Gaps 10;

OY 24 NYSAKDFSCAELTYTLEVOYKFEARAKMAA--ASGTVSSMFLYONGSEIADCRPWEVD 81
DB 40 slsldkssgsgfsgnqeflygkaevgmklvpgnsagtlvtlylk-----spgtwdeid 93

PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	22-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157863.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	28-OCT-1999;	99US-0162142.

	Matches	59;	Conservative	35;	Mismatches	97;	Indels	23;	Gaps	10;
OY	24	NVSAKDFGAGELYTEEYOYGKFERRIMKA--ASGYSSMFLYONGSEIADGRWVEYD	81	:	:	:	:	:	:	:
Dd	53	SLSLAKSGSgfsqnsqeflyjkaeqxlmLyvsnasgvlttffylk-----spgttwdeid	106	:	:	:	:	:	:	:
OY	82	IENVLK---NPSFQSNITTGKAGAOKNSEKHAAVPADAOFHHTYGLFWPNYYRWTPVD	138	:	:	:	:	:	:	:
Dd	107	felfllnsgfhyrltlhtnytt-kgsqdkeqgfllwdrlyvn--fhyceltmnpqrillftvd	163	:	:	:	:	:	:	:
OY	139	GOEVKRTEGGGVSLN-LTGTQLRF-NLMSSESAAMWG--QPDFSKLPLEFOQTINWKV	191	:	:	:	:	:	:	:
Dd	164	gpilrefkrneslgyprfpkpmrilyasaieweahatrglektdwskaprfayrynlv	223	:	:	:	:	:	:	:
OY	192	KYPTPGEGSGSDFTLLW-TDNFDFFDSRMCKG	224	:	:	:	:	:	:	:
Dd	224	egcvwanngksccpansswflqqlds-mqctlmkg	236	:	:	:	:	:	:	:

RESULT	13
AAgi18642	
ID	AAgi18642 standard; Protein; 298 AA.
XX	
AC	AAgi18642;
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 20133.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.
KW	
XX	
OS	Arabidopsis thaliana.
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PE	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.

PR 25-MAY-1999;	9905-0136021.
PR 27-MAY-1999;	9905-0136392.
PR 28-MAY-1999;	9905-0136782.
PR 01-JUN-1999;	9905-0137222.
PR 03-JUN-1999;	9905-0137528.
PR 04-JUN-1999;	9905-0137502.
PR 07-JUN-1999;	9905-0137724.
PR 08-JUN-1999;	9905-0138094.
PR 10-JUN-1999;	9905-0138540.
PR 14-JUN-1999;	9905-0138847.
PR 16-JUN-1999;	9905-0139119.
PR 16-JUN-1999;	9905-0139452.
PR 17-JUN-1999;	9905-0139493.
PR 18-JUN-1999;	9905-0139454.
PR 18-JUN-1999;	9905-0139455.
PR 18-JUN-1999;	9905-0139456.
PR 18-JUN-1999;	9905-0139457.
PR 18-JUN-1999;	9905-0139458.
PR 18-JUN-1999;	9905-0139459.
PR 18-JUN-1999;	9905-0139460.
PR 18-JUN-1999;	9905-0139461.
PR 18-JUN-1999;	9905-0139462.
PR 18-JUN-1999;	9905-0139463.
PR 18-JUN-1999;	9905-0139750.
PR 21-JUN-1999;	9905-0139817.
PR 22-JUN-1999;	9905-0139899.
PR 23-JUN-1999;	9905-0140353.
PR 23-JUN-1999;	9905-0140354.
PR 24-JUN-1999;	9905-0140695.
PR 28-JUN-1999;	9905-0140823.
PR 29-JUN-1999;	9905-0140991.
PR 30-JUN-1999;	9905-0141287.
PR 01-JUL-1999;	9905-0141842.
PR 01-JUL-1999;	9905-0142154.
PR 02-JUL-1999;	9905-0142055.
PR 06-JUL-1999;	9905-0142390.
PR 08-JUL-1999;	9905-0142803.
PR 09-JUL-1999;	9905-0142920.
PR 12-JUL-1999;	9905-0142977.
PR 13-JUL-1999;	9905-0143542.
PR 14-JUL-1999;	9905-0143624.
PR 15-JUL-1999;	9905-0144005.
PR 16-JUL-1999;	9905-0144085.
PR 16-JUL-1999;	9905-0144086.
PR 19-JUL-1999;	9905-0144325.
PR 19-JUL-1999;	9905-0144331.
PR 19-JUL-1999;	9905-0144332.
PR 19-JUL-1999;	9905-0144333.
PR 19-JUL-1999;	9905-0144334.
PR 19-JUL-1999;	9905-0144335.
PR 20-JUL-1999;	9905-0144352.
PR 20-JUL-1999;	9905-0144632.
PR 20-JUL-1999;	9905-0144884.
PR 21-JUL-1999;	9905-0144884.
PR 21-JUL-1999;	9905-0145086.
PR 21-JUL-1999;	9905-0145088.
PR 22-JUL-1999;	9905-0145085.
PR 22-JUL-1999;	9905-0145087.
PR 22-JUL-1999;	9905-0145089.
PR 22-JUL-1999;	9905-0145192.
PR 23-JUL-1999;	9905-0145145.
PR 23-JUL-1999;	9905-0145218.
PR 23-JUL-1999;	9905-0145224.
PR 26-JUL-1999;	9905-0145276.
PR 27-JUL-1999;	9905-0145913.
PR 27-JUL-1999;	9905-0145918.
PR 28-JUL-1999;	9905-0145951.
PR 02-AUG-1999;	9905-0146386.
PR 02-AUG-1999;	9905-0146388.
PR 02-AUG-1999;	9905-0146389.
<hr/>	
PR 03-AUG-1999;	9905-0147038.
PR 04-AUG-1999;	9905-0147204.
PR 05-AUG-1999;	9905-0147302.
PR 05-AUG-1999;	9905-0147192.
PR 05-AUG-1999;	9905-0147260.
PR 06-AUG-1999;	9905-0147303.
PR 06-AUG-1999;	9905-0147416.
PR 09-AUG-1999;	9905-0147493.
PR 09-AUG-1999;	9905-0147935.
PR 10-AUG-1999;	9905-0148171.
PR 11-AUG-1999;	9905-0148319.
PR 12-AUG-1999;	9905-0148341.
PR 13-AUG-1999;	9905-0148565.
PR 13-AUG-1999;	9905-0148684.
PR 16-AUG-1999;	9905-0149368.
PR 17-AUG-1999;	9905-0149175.
PR 18-AUG-1999;	9905-0149426.
PR 20-AUG-1999;	9905-0149722.
PR 20-AUG-1999;	9905-0149723.
PR 20-AUG-1999;	9905-0149929.
PR 23-AUG-1999;	9905-0149902.
PR 23-AUG-1999;	9905-0149930.
PR 25-AUG-1999;	9905-0150566.
PR 26-AUG-1999;	9905-0150884.
PR 27-AUG-1999;	9905-0151065.
PR 27-AUG-1999;	9905-0151066.
PR 30-AUG-1999;	9905-0151080.
PR 31-AUG-1999;	9905-0151303.
PR 01-SEP-1999;	9905-0151438.
PR 07-SEP-1999;	9905-0151930.
PR 10-SEP-1999;	9905-0152363.
PR 13-SEP-1999;	9905-0153070.
PR 15-SEP-1999;	9905-0153758.
PR 16-SEP-1999;	9905-0154018.
PR 20-SEP-1999;	9905-0154039.
PR 22-SEP-1999;	9905-0154779.
PR 23-SEP-1999;	9905-0155139.
PR 24-SEP-1999;	9905-0155486.
PR 24-SEP-1999;	9905-0155659.
PR 28-SEP-1999;	9905-0156458.
PR 29-SEP-1999;	9905-0156596.
PR 04-OCT-1999;	9905-0157117.
PR 05-OCT-1999;	9905-0157753.
PR 06-OCT-1999;	9905-0157865.
PR 07-OCT-1999;	9905-0158029.
PR 08-OCT-1999;	9905-0158232.
PR 12-OCT-1999;	9905-0158369.
PR 13-OCT-1999;	9905-0159293.
PR 13-OCT-1999;	9905-0159294.
PR 13-OCT-1999;	9905-0159295.
PR 14-OCT-1999;	9905-0159329.
PR 14-OCT-1999;	9905-0159330.
PR 14-OCT-1999;	9905-0159331.
PR 14-OCT-1999;	9905-0159637.
PR 14-OCT-1999;	9905-0159638.
PR 18-OCT-1999;	9905-0159584.
PR 21-OCT-1999;	9905-0160741.
PR 21-OCT-1999;	9905-0160767.
PR 21-OCT-1999;	9905-0160768.
PR 21-OCT-1999;	9905-0160770.
PR 21-OCT-1999;	9905-0160814.
PR 21-OCT-1999;	9905-0160815.
PR 22-OCT-1999;	9905-0160980.
PR 22-OCT-1999;	9905-0160981.
PR 22-OCT-1999;	9905-0160989.
PR 25-OCT-1999;	9905-0161404.
PR 25-OCT-1999;	9905-0161405.
PR 25-OCT-1999;	9905-0161406.
PR 26-OCT-1999;	9905-0161359.
PR 26-OCT-1999;	9905-0161360.
PR 26-OCT-1999;	9905-0161361.
PR 28-OCT-1999;	9905-0161920.
PR 28-OCT-1999;	9905-0161992.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136302.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142290.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145917.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158393.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:18:02 ; Search time 32.82 seconds

(without alignments)
619,702 Million bell updates/sec

Title: US-09-654-652a-2

Perfect score: 267

Sequence: 1 MSAKDFSGAEIYTLLEEVY.....PNSSSVDKLAALLENHHNNH 267

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	92.5	349	2 A44507	licheninase (EC 3.
2	8	3.0	381	2 S29560	fructose-bisphosph
3	8	3.0	407	2 T06408	probable fructose
4	8	3.0	429	2 T07815	S-locus-specific g
5	8	3.0	429	2 D70784	probable Rieske [2
6	7	2.6	60	2 C64698	probable histidine
7	7	2.6	95	2 S55439	hypothetical prote
8	7	2.6	115	2 P84469	probable glycine-r
9	7	2.6	115	2 H72583	hypothetical prote
10	7	2.6	119	2 S64079	probable membrane
11	7	2.6	139	2 T41242	hypothetical lysin
12	7	2.6	139	2 T41526	hypothetical prote
13	7	2.6	206	2 T25384	hypothetical prote
14	7	2.6	259	2 A70359	hydrogenase expres
15	7	2.6	270	2 A26480	knob protein - mal
16	7	2.6	304	2 F69864	hypothetical prote
17	7	2.6	306	2 A83551	still frameshift p
18	7	2.6	317	2 F82672	Arp sulfurylase, s
19	7	2.6	348	2 T08814	1,3-beta-glucanase
20	7	2.6	351	1 KGZOH1	histidine-rich gly
21	7	2.6	377	2 A32548	homeobox protein M
22	7	2.6	396	1 S15992	flavohemoglobin hm
23	7	2.6	396	2 F85900	hypothetical prote
24	7	2.6	420	2 T36193	probable salicylat
25	7	2.6	440	2 D64186	conserved hypotet
26	7	2.6	468	1 TVMSR2	transcription fact
27	7	2.6	469	1 TVHUE2	transcription fact
28	7	2.6	470	2 T20851	hypothetical prote
29	7	2.6	473	2 A54494	knob-associated hi

30	7	2.6	474	2 S75464	hypothetical prote
31	7	2.6	476	2 S57963	methyl Cpg binding
32	7	2.6	492	2 A41907	transcription fact
33	7	2.6	495	1 S31223	hypothetical prote
34	7	2.6	510	2 T20850	beta-galactosidase
35	7	2.6	546	2 B32688	conserved hypotet
36	7	2.6	611	2 F70325	knob protein precu
37	7	2.6	634	2 A54495	histidine-rich pro
38	7	2.6	637	2 A28412	probable fadD32 pr
39	7	2.6	637	2 E70887	knob-associated hi
40	7	2.6	654	2 B71623	knob-associated hi
41	7	2.6	657	2 A29454	lamin A - African
42	7	2.6	665	2 S02358	beta-galactosidase
43	7	2.6	677	2 A32611	hypothetical prote
44	7	2.6	715	2 H84799	hypothetical prote
45	7	2.6	955	2 T48515	hypothetical prote

ALIGNMENTS

RESULT 1
A44507
licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Teather, R.M.; Erfle, J.D.
J: Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1.3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-references: EMBL:M3676; NID:G148575; PIDN:AAA24896.1; PID:G148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 92.5%; Score 247; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.9e-244;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VSAKDFSGAEIYTLLEEVYGKFEARMKMAASGVTSMFYONGSEIADGRPMVEVDLEV 61
DB 25 VSAKDFSGAEIYTLLEEVYGKFEARMKMAASGVTSMFYONGSEIADGRPMVEVDLEV 84
QY 62 LKNGSFQSNITITGKAGAKTSEKHNHASPADAFTTYGLEWTPNYVVRTVDQEVYRK 121
DB 85 LKNGSFQSNITITGKAGAKTSEKHNHASPADAFTTYGLEWTPNYVVRTVDQEVYRK 144
QY 122 TEGGVSNLTGTGCLRFNLMSSSAAMYGQPDSEKLPYFQFNWVKYKYPGQEGGSD 181
DB 145 TEGGVSNLTGTGCLRFNLMSSSAAMYGQPDSEKLPYFQFNWVKYKYPGQEGGSD 204
QY 182 FTLDITDNFDFDGRSMGKGMTPFGNVRDLTDKNYSRDGLILALTRKQESFNGQVP 241
DB 205 FTLDITDNFDFDGRSMGKGMTPFGNVRDLTDKNYSRDGLILALTRKQESFNGQVP 264
QY 242 RDDEPAP 248
DB 265 RDDEPAP 271
RESULT 2
S29560
fructose-bisphosphatase (EC 3.1.3.11) - garden pea (fragment)
C:Species: Pisum sativum (garden pea)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S29560
R:Carasco, J.L.; Chueca, A.; Hermoso, R.; Lazaro, J.J.; Ramos, J.L.; Sahrawy, M.; Pr
submitted to the EMBL Data Library, October 1992
A:Description: Cloning, structure and expression of a pea cDNA clone coding for a pho

A:Reference number: S29560
A:Accession: S29560
A:Molecule type: mRNA
A:Residues: 1-381 <CAR>
A:Cross-references: EMBL:X68826; NID:g20716; PIDN:CAA48719.1; PID:g20717
C:Superfamily: fructose-bisphosphatase
C:Keywords: phosphoric monoester hydrolase

Query Match 3.0%; Score 8; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 SMLGTG 135
|||||||
DB 86 SMLGTG 93

RESULT 3
T06408
Probable fructose-bisphosphatase (EC 3.1.3.11) precursor - garden pea chloroplast
C:Species: chloroplast *Plum sativum* (garden pea)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Jun-1999
C:Accession: T06408
R:Hahn, T.R.; Dong, S.M.; Rhim, J.H.
Submitted to the EMBL Data Library, February 1999
A:Description: cDNA sequence and red light modulated expression of chloroplast fructose-
A:Reference number: Z15655
A:Accession: T06408
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: translated from GB/EMBL/DBJ
A:Residues: 1-407 <IAH>
A:Cross-references: EMBL:L24806; NID:g609560; PIDN:AD10213.1; PID:g609561
A:Experimental source: cv. Giant
C:Genetics:
A:Genome: chloroplast
C:Superfamily: fructose-bisphosphatase
C:Keywords: chloroplast; phosphoric monoester hydrolase
F:1-50/Domain1: transit peptide (chloroplast) #status predicted <TNP>
F:51-407/Product: fructose-bisphosphatase #status predicted <MAN>

Query Match 3.0%; Score 8; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 SMLGTG 135
|||||||
DB 112 SMLGTG 119

RESULT 4
T07815
S-locus-specific glycoprotein S8 - radish (fragment)
C:Species: *Raphanus sativus* (radish)
C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07815
R:Sakamoto, K.; Kusaba, M.; Nishio, T.
Mol. Gen. Genet. 258, 397-403, 1998
A:Title: Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gen
A:Reference number: Z16146; MUID:98311079
A:Accession: T07815
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-429 <SAK>
A:Cross-references: EMBL:AB009683; NID:g3327851; PIDN:BAA31730.1; PID:g3327852
C:Genetics:
A:Gene: SLG(S8)
C:Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology
C:Keywords: glycoprotein
F:32-427/Domain: S-locus-specific glycoprotein homology <SSG>

Query Match 3.0%; Score 8; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 213 TDKNYSR 220
|||||||
DB 253 TDKNYSR 260

RESULT 5
D70784
Probable Rleske [2Fe-2S] iron-sulfur protein cyoB - Mycobacterium tuberculosis (strai
C:Species: *Mycobacterium tuberculosis*
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70784
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtrold,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: D70784
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-429 <COL>
A:Cross-references: GB:Z70283; GB:AL123456; NID:g3261561; PIDN:CAA94264.1; PID:e23357
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: qcrA
C:Superfamily: Rleske [2Fe-2S] homology
C:Keywords: 2Fe-2S; metalloprotein; Rleske iron-sulfur protein
F:343-389/Domain: Rleske [2Fe-2S] homology <RSK>
F:353.355.372.375/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #statu

Query Match 3.0%; Score 8; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 RKGESFN 237
|||||||
DB 335 RKGESFN 342

RESULT 6
C64698
Probable histidine-rich metal-binding protein - *Helicobacter pylori*
C:Species: *Helicobacter pylori*
A:Variety: strains J99, 26695
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: C64698; C71821
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaf, H.G.; Glöck, A.; McKe
non, J.D.; Kelley, J.M.; Cotton, M.D.; Feldman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karik, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A:Reference number: A64520; MUID:97934467
A:Accession: C64698
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-60 <TOM>
A:Cross-references: GB:AE000643; GB:AE000511; NID:g2314598; PIDN:AA008471.1; PID:g231
A:Experimental source: strain 26695
R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, O.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: C71821
A:Molecule type: DNA
A:Residues: 1-60 <ARN>

A:Cross-references: GB:AE001555; GB:AE001439; NID:g4155929; PIDN:AAD06898.1; PID:g415593
A:Experimental source: strain J99
C:Genetics:
A:Gene: HP1427; jhp1320

Query Match 2.6%; Score 7; DB 2; Length 60;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHNNH 267
DB 27 EHHNNH 33

RESULT 7
S55439
hypothetical protein ywkr - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
C:Accession: S55439; E70061
R:Glaser, P.; Danchin, A.
submitted to the EMBL Data Library, May 1995
A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region from 3
A:Reference number: S55414
A:Accession: S55439
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-95 <GLA>
A:Cross-references: EMBL:Z49782; NID:9853752; PIDN:CA69886.1; PID:9853778
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
leeh, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
R.; Koenigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogilawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
Aechul, M.; Tamakoshi, A.; Tanaka, T.; Terpiltra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: E70061
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-95 <KUN>
A:Cross-references: GB:Z99122; GB:AL009126; NID:g2636029; PIDN:CA615716.1; PID:e1184605;
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywkr

Query Match 2.6%; Score 7; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKKMAA 32
DB 71 RKKMAA 77

RESULT 8
F84469
probable glycine-rich protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84469
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umayam, L.; Tallon, L.

euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: F84469
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-115 <STO>
A:Cross-references: GB:AE002093; NID:g4581171; PIDN:AAD24654.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g05530
A:Map position: 2

Query Match 2.6%; Score 7; DB 2; Length 115;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 AAASGV 36
DB 22 AAASGV 28

RESULT 9
H72583
hypothetical protein APE1948 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H72583
R:Kavayalavasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
dna, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: H72583
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-115 <KAW>
A:Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80957.1; PID:d1044743; PID:g
C:Genetics:
A:Experimental source: strain K1
A:Gene: APE1948

Query Match 2.6%; Score 7; DB 2; Length 115;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHNNH 267
DB 31 EHHNNH 37

RESULT 10
S64079
probable membrane protein YGL072c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G3263
C:Species: Saccharomyces cerevisiae
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jun-2000
C:Accession: S64079
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64071
A:Accession: S64079
A:Molecule type: DNA
A:Residues: 1-119 <RIE>
A:Cross-references: EMBL:Z72596; NID:g1322584; PID:e243291; PID:g1322587; GSPDB:GN000
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YGL072c
A:Map position: 7L
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YGL072c
C:Keywords: transmembrane protein

F:35-51/Domain: transmembrane #status predicted <TM>

Query Match 2.6%: Score 7; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 FSGAELY 13
|||||
DB 98 FSGAELY 104

RESULT 11

T41242

hypothetical lysine-rich protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T41242

R:Lyne, M.; Brown, S.; Quail, M.; Harris, D.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, July 1999

A:Reference number: 221980

A:Accession: T41242

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-139 <LYN>

A:Cross-references: EMBL:AL109608; PIDN:CAB51561.1; GSPDB:GN00068; SPDB:SPCC23B6.02C

A:Experimental source: strain 972h-; cosmid c23B6

C:Genetics:

A:Gene: SPDB:SPCC23B6.02C

A:Map position: 3

A:Introns: 106/1

Query Match 2.6%: Score 7; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 254 DKLAAL 260
|||||
DB 74 DKLAAL 80

RESULT 12

T41526

hypothetical protein SPCC645.09 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T41526

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.

submitted to the EMBL Data Library, March 1999

A:Reference number: 222000

A:Accession: T41526

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-139 <MOO>

A:Cross-references: EMBL:AL049498; PIDN:CAB39905.1; GSPDB:GN00068; SPDB:SPCC645.09

A:Experimental source: strain 972h-; cosmid c645

C:Genetics:

A:Gene: SPDB:SPCC645.09

A:Map position: 3

Query Match 2.6%: Score 7; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 244 DEPAAPS 250
|||||
DB 112 DEPAAPS 118

RESULT 13

T25384

hypothetical protein T27F6.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T25384

R:Johnson, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: 220026

A:Accession: T25384

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-206 <RTL>

A:Cross-references: EMBL:Z82060; PIDN:CAB04883.1; GSPDB:GN00019; CESP:T27F6.4

A:Experimental source: clone T27F6

C:Genetics:

A:Gene: CESP:T27F6.4

A:Map position: 1

A:Introns: 61/3; 154/3

Query Match 2.6%: Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHNNHH 267
|||||
DB 115 EHHNNHH 121

RESULT 14

A70359

hydrogenase expression/formation protein B - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 24-Sep-1999

C:Accession: A70359

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: A70359

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-259 <AQF>

A:Cross-references: GB:AE000701; NID:g2983260; PIDN:AAC06860.1; PID:g2983266; GB:AE00

C:Experimental source: strain VF5

C:Genetics:

A:Gene: hypB

C:Superfamily: hydrogenase expression/formation protein hypB

Query Match 2.6%: Score 7; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHNNHH 267
|||||
DB 17 EHHNNHH 23

RESULT 15

A26480

knob protein - malaria parasite (Plasmodium falciparum) (fragments)

C:Species: Plasmodium falciparum

C:Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 09-Jun-2000

C:Accession: A26480

R:Kilejian, A.; Sharma, Y.D.; Karoui, H.; Naslund, L.

Proc. Natl. Acad. Sci. U.S.A. 83, 7938-7941, 1986

A:Title: Histidine-rich domain of the knob protein of the human malaria parasite Plas

A:Reference number: A26480; MUID:87017062

A:Accession: A26480

A:Molecule type: mRNA

A:Residues: 1-270 <KIL>

A:Cross-references: GB:M14210; NID:g160361; PIDN:AAA29631.1; PID:g160362

C:Superfamily: knob-associated histidine-rich protein

Query Match 2.6%; Score 7; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHHHH 267
|||||
Db 62 EHHHHH 68

Search completed: August 23, 2001, 11:18:03
Job time: 210 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:15:07; Search time 32.87 Seconds

(without alignments)
618.759 Million cell updates/sec

Title: US-09-654-652a-2

Perfect score: 1439

Sequence: 1 MVSADFGAELTYLEEVY.....PNSSVDKLAALAEHHHHH 267

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.68:*
2: pir1:*
3: pir2:*
4: pir3:*
5: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	93.1	349	2	A44507
2	266.5	18.5	851	2	H84053
3	265.5	18.5	252	2	A48378
4	247.5	17.2	276	2	140453
5	245.5	17.1	242	1	LXBS
6	242.5	16.9	802	2	A36910
7	238.5	16.6	239	1	A29091
8	237	16.5	334	1	S23498
9	229.5	15.9	243	1	S15388
10	226	15.7	237	1	S11927
11	225	15.6	238	1	S19012
12	222.5	15.5	242	2	J50611
13	190.5	13.2	302	2	G84053
14	182.5	12.7	507	2	S64507
15	180	12.5	268	2	S34804
16	172	12.0	282	2	T02354
17	171.5	11.9	277	2	B83354
18	169.5	11.8	282	2	A85354
19	169	11.7	642	2	B72428
20	163	11.3	286	2	S71225
21	159	11.0	287	2	T04236
22	158.5	10.7	277	2	S71222
23	150.5	10.5	269	2	S61555
24	147	10.2	467	2	S30839
25	146.5	10.2	284	2	T52097
26	142	9.9	422	2	S48564
27	141.5	9.8	289	2	T06166
28	140.5	9.8	280	2	T02090
29	140.5	9.8	310	2	A86239

30	140	9.7	305	2	G84568	probable xyloglucan
31	139.5	9.7	286	2	T06202	xyloglucan endo-1,
32	133.5	9.3	286	2	S48201	licheninase (EC 3,
33	132	9.2	295	2	T10210	xyloglucan endo-1,
34	131.5	9.1	292	2	T06201	xyloglucan endo-1,
35	129.5	9.0	1324	2	T18265	endo-1,3(4)-beta-g
36	128	8.9	283	2	T07678	xyloglucan endo-1,
37	127.5	8.9	269	2	T05895	xyloglucan endo-1,
38	127	8.8	292	2	T04514	xyloglucan endo-1,
39	125.5	8.7	277	2	T01566	end-xyloglucan tra
40	124	8.6	299	2	D84519	probable endoxylol
41	123	8.5	296	2	C49539	xyloglucan endo-1,
42	121.5	8.4	292	2	T10211	xyloglucan endo-1,
43	121	8.4	284	2	T06200	xyloglucan endo-1,
44	121	8.4	293	2	T10523	xyloglucan endo-1,
45	120	8.3	877	2	JN0772	glucan endo-1,3-be

ALIGNMENTS

RESULT 1
A44507
Licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Teather, R.M.; Erfle, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1,3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-references: EMBL:M3676; NID:g148575; PIDD:AAA24896.1; PID:g148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 93.1%; Score 1340; DB 2; Length 349;
Best Local Similarity 99.6%; Pred. No. 7.8e-105;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VSAKDFSGAELTYLEEVYQKFEARMKMAASGVSSMFLYONGSEIADGRPWVEDIEV 61
DB 25 VSAKDFSGAELTYLEEVYQKFEARMKMAASGVSSMFLYONGSEIADGRPWVEDIEV 84
QY 62 LCKNPGSFQSNITITGKAGAKTSEKHNHVASPAADQAFHTYGLEWTPNVRWTVDGQEVK 121
DB 85 LCKNPGSFQSNITITGKAGAKTSEKHNHVASPAADQAFHTYGLEWTPNVRWTVDGQEVK 144
QY 122 TEGGQVSNLTGQGLRFNLMSSSAAWGFQFDESKLPLOFNNMVVVVYVTGGGEGSSD 181
DB 145 TEGGQVSNLTGQGLRFNLMSSSAAWGFQFDESKLPLOFNNMVVVVYVTGGGEGSSD 204
QY 182 FTLDWTDNFDTFDGSFWMGKDMTFDGNRYDLTDKNISRDGMLLALTRKGQESFNGQVP 241
DB 205 FTLDWTDNFDTFDGSFWMGKDMTFDGNRYDLTDKNISRDGMLLALTRKGQESFNGQVP 264
QY 242 RDDEPAPNSS 252
DB 265 RDDEPAPQSSS 275

RESULT 2
H84053
endo-beta-1,3-1,4 glucanase (Licheninase) bgis [imported] - Bacillus halodurans (stra
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
C:Accession: H84053
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Molecule type: DNA
A:Residues: 1-242 <KUN>
A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CA815943.1; PID:g2636453
R:Experimental source: strain 168
R:Murphy, N.; McConnell, D.J.; Cantwell, B.A.
Nucleic Acids Res. 12, 5355-5367, 1984
A>Title: The DNA sequence of the gene and genetic control sites for the excreted B. subtilis
A:Reference number: A93526; MUID:04272222
A:Accession: A22914
A:Molecule type: DNA
A:Residues: 1-203,'L',205-242 <MUR>
A:Cross-references: EMBL:X00754; NID:g39818; PIDN:CAA25328.1; PID:g685236
A:Experimental source: strain C120
A>Note: the authors translated the codon CAA for residue 29 as Lys and CCA for residue 8
R:Tezuka, H.; Yunkai, T.; Yabuuchi, S.
Agric. Biol. Chem. 53, 2335-2339, 1989
A>Title: Construction of a beta-glucanase hyperproducing *Bacillus subtilis* using the old
A:Reference number: A90026
A:Accession: A90026
A:Molecule type: DNA
A:Residues: 1-23,'S',25-82,'S',84-242 <TEZ>
A:Cross-references: DBJ:D00518; NID:g216243; PIDN:BAA00405.1; PID:g216244
A:Experimental source: strain Y-25, clone pLE100
R:Yunkai, T.; Tezuka, H.; Yabuuchi, S.
Agric. Biol. Chem. 53, 2341-2346, 1989
A>Title: Purification and some properties of two enzymes from a beta-glucanase hyperprod
A:Reference number: A90027
A:Contents: annotation
A>Note: source was hyperproducing strain HL-25 with gene from strain Y-25
A>Note: the amino ends of the mature forms of E-1 and E-2 are pyroglutamic acid and glut
C:Genetics:
A:Gene: bglIS
C:Function:
A>Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenan and c
C:Superfamily: licheninase
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation; P
F:1-28/Domin: signal sequence #status predicted <SIG>
F:29-24/Product: lichenanase #status predicted <MAT>
F:29/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) (partial) #status

Query Match 17.1% Score 245.5; DB 1; Length 242;
Best Local Similarity 35.8%; Pred. No. 2.6e-13;
Matches 59; Conservative 24; Mismatches 71; Indels 11; Gaps 6;

OY 7 FSGALYTLIEVQYQGFAPRMKAASGTVSMPLYONGSEIADGRPVVEVDIELGNP 66
| : : : | : | | | | : | | | | : | : | : | : | : | :
Db 87 FDCCGNRSRVQYGYGLYEVRMKPAKNIGIVSSFITYPT---DGPMDIEDIEFLGDT 143
| : : : | : | | | | : | : | : | : | : | : | : | : | :
OY 67 GSPOSNIITGAKAGOKTISEKHNAVSPAADOAFHTYGLEMTFNVMYTVDGGEVKTEGG 126
| : : : | : | | | | : | : | : | : | : | : | : | : | :
Db 144 TKVOFNYYTNAG--NNEKITVDLGFDANAYHGAFFDMOPNSIKWYDQG-LKHTATNQ 199
| : : : | : | | | | : | : | : | : | : | : | : | : | :
OY 127 VSNLTGTGLRFNLMSSEA-AWVGQPFDESKLPLEQFINMWKYK 170
| : : : | : | | | | : | : | : | : | : | : | : | : | :
Db 200 IPTTPGK--IMNNLMNGTGVDMLGSYGVN-PLYAHDWRYTK 241
| : : : | : | | | | : | : | : | : | : | : | : | : | :

RESULT 6
A36910
xylanase, beta(1,3-1,4)-glucanase - Ruminococcus flavefaciens
C:Species: Ruminococcus flavefaciens
C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
R:Accession: A36910
R:Flint, H.J.; Martlin, J.; McPherson, C.A.; Daniel, A.S.; Zhang, J.X.
J. Bacteriol. 175, 2643-2651, 1993
A>Title: A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-glucanase domat
A:Reference number: A36910; MUID:93259938
A:Accession: A36910
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <FLT>
A:Cross-references: GB:S61204; NID:g385910; PIDN:AAB26620.1; PID:g385911

C:Note: sequence extracted from NCBI backbone (NCBI:131871, NCBI:131872)
 C:Superfamily: endo-1,4-beta-xylanase homology; Thermotoga xylanase A amino-terminal
 F:42-235/Domain: endo-1,4-beta-xylanase homology <YLL>
 F:259-401/Domain: Thermotoga xylanase A amino-terminal repeat homology <YXA>

```

Query Match          16.9%  Score 242.5;  DB 2;  Length 802;
Best Local Similarity 34.9%;  Pred. No. 2,1e-12;
Matches 58;  Conservative 21;  Mismatches 74;  Indels 13;  Gaps 6;

OY 7 FSGAELTYLEVOYQGFKEFARKMAAASGTVSSMFLYONGSEIADGRPVEVDIEVLGKNP 66
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 634 YSGGEFRINNNHYHYGYSCMQAMKNDGVSSFFTYGCP---DDNPWDEIDIEILGKNT 690
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 67 GSPFSNITTGKAGAKQKTSSEKHHAVSPAADQAFHTYGTLEMTNRYKRVTVDOGVKTEGGQ 126
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 TVQGVNYYTNGCGKH---EKLYDLGFDSSSEAYHYTGFDQPCNRYIAMVYDGREYVRA--- 743
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 127 VSNLTGTGQ-LRFNLMSSESA-AWVGQFDESKLPLFOFINNVKYYK 170
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 744 TQDIPKTPGKIMNMAWPGLTIVDWMLKAFN-GRPTLTAHYQWVYVYNNK 788
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
A29091
Licheninase (EC 3.2.1.73) beta - Bacillus amyloliquefaciens
N:Alternate names: 1,3-1,4-beta-D-glucan 4-glucanohydrolase; beta-glucanase; lichenase
C:Species: Bacillus amyloliquefaciens
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A29091
R:Hotemeler, J.; Kurtz, A.; Borriess, R.; Knowles, J.
Gene 49, 177-187, 1986
A:Title: The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology
A:Reference number: A91564; MUID:87192007
A:Accession: A29091
A:Molecule type: DNA
A:Residues: 1,239 <HOF>
A:Cross-references: GB:M15674; NID:g143009; PIDN:AAA87323.1; PID:g143010
A:Experimental source: strain BE20/78
C:Genetics:
A:Gene: bglA
C:Superfamily: Licheninase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match          16.6%  Score 238.5;  DB 1;  Length 239;
Best Local Similarity 34.3%;  Pred. No. 9,8e-13;
Matches 58;  Conservative 26;  Mismatches 74;  Indels 11;  Gaps 6;

OY 3 SAKDFSGAELTYLEVOYQGFKEFARKMAAASGTVSSMFLYONGSEIADGRPVEVDIEVL 62
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 SYNKFDGCGENSVQYGYGLVEYRMRPAKNGIVSSFFTYGTPTE---GTPWDEIDIEFL 136
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 63 GKNPSSPFSNITTGKAGAKQKTSSEKHHAVSPAADQAFHTYGTLEMTNRYKRVTVDOGEVAKT 122
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 GKDTTKVQFNYYTNGAG---NHEKPADLGFDPAANAHYHVAADWOPNSITKMYVDQ-LKRT 192
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 123 EGGGVSNLTGTQGLRFLNLMSSESA-AWVGQFDESKLPLFOFINNVKYYK 170
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 ATTQIPAAAPGK--IMNMLNNGTGVDDWLGSTYNGVN-PIYAHYDWRKRYK 238
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
S23498
Licheninase (EC 3.2.1.73) lichen precursor - Clostridium thermocellum
N:Alternate names: beta-1,3-1,4-glucanase lichen; lichenase lichen
C:Species: Clostridium thermocellum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S23498; S22137
R:Schlimming, S.; Schwarz, W.H.; Staudenbauer, W.L.
Eur. J. Biochem. 204, 13-19, 1992
A:Title: Structure of the Clostridium thermocellum gene lichen and the encoded beta-1,3
    1 cellulases.
  
```


A:Cross-references: EMBL:X57094; NID:g48815; PIDN:CAA0379.1; PID:g48817
 C:Function:
 A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in Lichenin and C:
 C:Superfamily: Licheninase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-26/Domain: signal sequence #status predicted <PRO>
 F:27-237/Product: Licheninase #status predicted <MAY>
 F:56-85/Disulfide bonds: #status predicted
 F:129/Active site: Glu #status predicted

Query Match 15.6%, Score 225; DB 1; Length 238;
 Best Local Similarity 34.5%; Pred. No. 1.3e-11;
 Matches 58; Conservative 23; Mismatches 71; Indels 16; Gaps 8;

OY 7 FSGAELTYLEEVQYKFEARKMAAAGTSSMFLYONGSEIADGRPWEVDIEVLGNP 66
 DB 83 FDGGEYRSTNNYGLYEVSMKPAKNTGIVSSFFTYTGPSSH--GTQWDEIDIEFLGKDT 139
 OY 67 GSFGSNITTKAGAAQKTESEKHHAVSPADQAFHTYGLEMTPTNYRWYVDGGEVRKTEGQ 126
 DB 140 TKVQFNYYTNGVGH--EKIINLGFDASTSHYTAFFDMQPCYIKMYVDG-VLKHT---A 192
 OY 127 VSNLTGTGQ-LRFNLMSSESA-AWVGQFDESKLPLEQFINWVKYKYP 172
 DB 193 TTNISTPTGKIMMNLNMGTVDSWLSYNGAN-PLYAEYDWV---KYT 236

RESULT 12
 JS0611
 endo-1,3(4)-beta-glucanase (EC 3.2.1.6) precursor - Clostridium thermocellum
 N:Alternate names: laminarinase
 C:Species: Clostridium thermocellum
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jan-1997
 C:Accession: JS0611; S18726
 R:Zverlov, V.V.; Lapev, D.A.; Tishkov, V.I.; Velikodvorskaja, G.A.
 Biochem. Biophys. Res. Commun. 181, 507-512, 1991
 A:Title: Nucleotide sequence of the Clostridium thermocellum laminarinase gene.
 A:Reference number: JS0611; MUID:92095946
 A:Accession: JS0611

A:Molecule type: DNA
 A:Residues: 1-242 <ZVEI>
 A:Cross-references: EMBL:X58392
 R:Zverlov, V.V.; Velikodvorskaja, G.A.
 Biotechnol. Lett. 12, 811-816, 1990
 A:Title: Cloning the Clostridium thermocellum thermostable laminarinase gene in Escheric
 A:Reference number: S18726
 A:Accession: S18726

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-242 <ZVE2>
 A:Cross-references: EMBL:X58392
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1991
 C:Genetics:
 A:Gene: lam1
 C:Superfamily: Licheninase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-242/Product: endo-1,3(4)-beta-glucanase #status predicted <END>

Query Match 15.5%; Score 222.5; DB 2; Length 242;
 Best Local Similarity 35.6%; Pred. No. 2.2e-11;
 Matches 58; Conservative 22; Mismatches 64; Indels 19; Gaps 7;

OY 5 KDFSGAELTYLEEVQ-----YKFEARKMAAAGTSSMFLYONGSEIADGRPWEVD 58
 DB 80 REYGSSTYKSGEYRTKGFPGYGYEVKMAKKNIGIVSSFFTYTGPSS--DNNPWEID 136
 OY 59 IEVLGNFSGQSNITTKAGAAQKTESEKHHAVSPADQAFHTYGLEMTPTNYRWYVDGQ 118
 DB 137 IEFLGKDTTKVQFNMWYKGVG---NEYILNLGFDASODFHTYGFEMRPDYIDFVVDGK 193

OY 119 VRKTEGGVSNLTGTGQ-LRFNLMSSESA-AWVGQFDESKLP 159
 DB 194 VYR----GTRNIPYTPRKIMMNLMPGIGVDEMLGRYD-GRPPL 231

RESULT 13
 G84053
 hybrid-endo-beta-1,3-1,4 glucanase BH3231 [Imported] - Bacillus halodurans (strain C-
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: G84053
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20263314

A:Accession: G84053
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-302 <STG>
 A:Cross-references: GB:AB001518; GB:BA000004; NID:g10175792; PIDN:BA06950.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3231

Query Match 13.2%; Score 190.5; DB 2; Length 302;
 Best Local Similarity 27.8%; Pred. No. 1.4e-08;
 Matches 62; Conservative 34; Mismatches 98; Indels 29; Gaps 9;

OY 6 DFSGAELTYLEEVQYKFEARKMAAAGTSSMFLYONGSEIADGRPWEVDIEVLGN 65
 DB 16 NECAEYSTNHFHYGLYEVSMKPSNVSGVSSFFTYTGPSS--YNGAPWDEIDIEFLGND 73
 OY 66 PSFGSNITTKAGAAQKTESEKHHAVSPADQAFHTYGLEMTPTNYRWYVDGGEVRKTEG 125
 DB 74 TKVQFNYYTNGVGH--EKIINLGFDASTSHYTAFFDMQPCYIKMYVDG-VLKHT---A 192
 OY 126 QVSNLTGTGQ-LRFNLMSSESA-AWVGQFDESKLPLEQFINWVKYKYP---GCG 176
 DB 131 IPSN---PSKMMNINWNTYTGIDEMAGATWGNAN\$-----YEW---RTFPNNEBSOT 178
 OY 177 EGSDFTLDWTDNFTFDGSRN--GKGDWTFPDGNEVLDLTKNI 217
 DB 179 PIASDFQLHACEYSDARGVTSWQCGVGSF-YPGSMWKFDNVNL 220

RESULT 14
 S64507
 probable membrane protein YGR189C - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein G7553
 C:Species: Saccharomyces cerevisiae
 C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 29-Oct-1999
 C:Accession: S64507
 R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64499

A:Accession: S64507
 A:Molecule type: DNA
 A:Residues: 1-507 <ARP>
 A:Cross-references: EMBL:Z72974; NID:g1323335; PID:e243566; PID:g1323336; GSPDB:GN000
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: MIPS:YGR189C
 A:Map position: 7R
 C:Keywords: transmembrane protein
 F:6-22/Domain: transmembrane #status predicted <TM>

Query Match 12.7%; Score 182.5; DB 2; Length 507;
 Best Local Similarity 26.7%; Pred. No. 1.2e-07;
 Matches 58; Conservative 37; Mismatches 91; Indels 31; Gaps 8;

OY 1 MVSADFGAELTYLEEVQYKFEARKMAAAGTSSMFLYONGSEIADGRPWEVDIE 60

```

Db      87  MTLAKRRDNPGLKSNFYIMCKLEVIILKANGTIVSSFYLQSDLDL-----EIDIE 138
QY      61  VLGNPQSFQSNITITGKAGAKTSEKHHAVPADQAFHTYGLKWTPTVYRWTVGQGEVR 120
Db      139  WVGDNTPQFQSNFESKGDITTYDRGEFHGVDPYDK-FHNTLLDAMDKTWTYLDGESVR 197
QY      121  KTEGGQVSNLTGTG-----LRFNLMS--ESAA---WV-GQFDESKLPLFOFINW 165
Db      198  -----VLSNTSSEGYQSPMYLMMGIWAGGDPNAGTIEMAGETINYNDAFTMYIEK 251
QY      166  VKYKTYTPGQCEGSDFTLMDTDFDTSRMRGKCD 202
Db      252  VIVTDYSTGKKYTYGDQGSW-ESIEADGGSIVGRYD 287

```

RESULT 15

```

S34804
exok protein - Rhizobium meliloti megaplasmid 2
C:Species: Rhizobium meliloti
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: S34804; S30072
R:Becker, A.; Kleichmann, A.; Arnold, W.; Puhler, A.
Mol. Gen. Genet. 238, 145-154, 1993
A>Title: Analysis of the Rhizobium meliloti exoX/exoK/exoL fragment: ExoK shows homology
A:Reference number: S34803; MUID:93241147
A:Accession: S34804
A:Molecule type: DNA
A:Residues: 1-268 <BEC>
A:Cross-references: EMBL:Z17219
A>Note: The authors translated the codon GTA for residue 208 as Leu
C:Genetics:
A:Gene: exok
A:Genome: plasmid
C:Superfamily: licheninase

```

```

Query Match      12.5%; Score 180; DB 2; Length 268;
Best Local Similarity 33.8%; Pred. No. 9e-08;
Matches 46; Conservative 26; Mismatches 52; Indels 12; Gaps 6;

```

```

QY      20  YGKFEARMKMAASGIVSSFLYONGSEIADGRPWVEVDIEVLGNPQSFQSN-IITGKA 78
Db      104  YGYEARIKADSGSLNSAFFTYIGP--ADKKPHDEIDEEVLGNKTAQVQINQVVSAGK 160
QY      79  GAQKTSKHHAVSPADQAFHTYGLKWTPTVYRWTVGQGEVRK-TEGGQVSNLTGTGLR 137
Db      161  GNEFLAD----VEGGANQGENDYAFVVEKKNRIRIYVNGELVHEVTDPAKIP--VNAQKIF 214
QY      138  FNLMSSESA-WVGQF 152
Db      215  FSLMGFTDLTDMWGTF 230

```

Search completed: August 23, 2001, 11:15:07
 Job time: 94 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:15:07 ; Search time 32.87 Seconds
(without alignments)
808.790 Million cell updates/sec

Title: US-09-654-652a-3
Perfect score: 1824
Sequence: 1 MNIKRTAVKSAALAVAAAAA.....AKGAKVNPNGHKRYVNFPH 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1824	100.0	349	A44507	licheninase (EC 3. endo-beta-1,3-1,4
2	268.5	14.7	851	H84053	licheninase (EC 3. licheninase (EC 3.
3	265.5	14.6	252	A48378	licheninase (EC 3. licheninase (EC 3.
4	248	13.6	242	1A85	licheninase (EC 3. licheninase (EC 3.
5	247.5	13.6	276	140453	licheninase (EC 3. licheninase (EC 3.
6	242.5	13.3	239	A29091	licheninase (EC 3. xylanase, beta(1,3
7	242.5	13.3	802	A36910	licheninase (EC 3. licheninase (EC 3.
8	237	13.0	334	S23498	licheninase (EC 3. licheninase (EC 3.
9	233.5	12.8	243	S15388	licheninase (EC 3. probable membrane
10	230	12.6	507	S64507	licheninase (EC 3. licheninase (EC 3.
11	228	12.5	237	S11927	licheninase (EC 3. endo-1,3(4)-beta-g
12	227.5	12.5	238	S19012	licheninase (EC 3. hybrid-endo-beta-1
13	224.5	12.3	242	J50611	licheninase (EC 3. UTR2 protein - yea
14	190.5	10.4	302	G84053	endo-1,3(4)-beta-1
15	189.5	10.4	467	S30839	endo-1,3(4)-beta-1
16	180.5	9.9	268	S24804	endo-1,3(4)-beta-1
17	173	9.5	282	T02354	endo-1,3(4)-beta-1
18	173	9.5	642	B72428	endo-1,3(4)-beta-1
19	172.5	9.5	277	B83554	endo-1,3(4)-beta-1
20	170.5	9.3	282	A85354	endo-1,3(4)-beta-1
21	164	9.0	286	S71225	endo-1,3(4)-beta-1
22	159.5	8.7	287	T04236	endo-1,3(4)-beta-1
23	157.5	8.6	269	S61555	endo-1,3(4)-beta-1
24	154.5	8.5	277	S71222	endo-1,3(4)-beta-1
25	149	8.2	289	T06166	endo-1,3(4)-beta-1
26	146.5	8.0	284	T52097	endo-1,3(4)-beta-1
27	145.5	8.0	310	A86239	endo-1,3(4)-beta-1
28	145	7.9	305	G84568	endo-1,3(4)-beta-1
29	140.5	7.7	280	T02090	endo-1,3(4)-beta-1

30	139.5	7.6	286	2	T06202	xyloglucan endo-1,
31	138	7.6	422	2	S48564	probable membrane
32	137	7.5	292	2	T06201	xyloglucan endo-1,
33	133.5	7.3	286	2	S48201	licheninase (EC 3.
34	132	7.2	295	2	T10210	xyloglucan endo-1,
35	128.5	7.1	1324	2	T18265	endo-1,3(4)-beta-g
36	128.5	7.0	277	2	J60156	endo-1,3(4)-beta-g
37	128.5	7.0	283	2	T04514	endo-xyloglucan tra
38	128	7.0	283	2	T07678	xyloglucan endo-1,
39	127.5	7.0	269	2	T05895	xyloglucan endo-1,
40	127	7.0	284	2	T06200	xyloglucan endo-1,
41	125.5	6.9	299	2	D84519	probable endoxylig
42	123	6.7	296	2	C49539	xyloglucan endo-1,
43	121.5	6.7	292	2	T10211	xyloglucan endo-1,
44	121	6.6	289	2	F71402	xyloglucan endo-1,
45	121	6.6	293	2	T10523	xyloglucan endo-1,

ALIGNMENTS

RESULT 1
A44507
licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Teather, R.M.; Erfle, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1,3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-references: EMBL:M33676; NID:9148575; PIDD:AAA24896.1; PID:9148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 100.0% Score 1824; DB 2; Length 349;
Best Local Similarity 100.0% Pred. No. 7.6e-123;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNIKRTAVKSAALAVAAAAAALTTNYSAKDFSGAEELTYEEVYGGFKFARMKMAASGTVS 60
DB 1 MNIKRTAVKSAALAVAAAAAALTTNYSAKDFSGAEELTYEEVYGGFKFARMKMAASGTVS 60
OY 61 SMFLYONGSEIADGRPWVEVDIEVLGKNPGSFOSNITGKAGAOQTSEKHAHVSPADQA 120
DB 61 SMFLYONGSEIADGRPWVEVDIEVLGKNPGSFOSNITGKAGAOQTSEKHAHVSPADQA 120
OY 121 PFTYGLWTPNVPVMTVGOEVRKTEGGQVSNLTGTGRLFMSSSAAMVGOPEDESKL 180
DB 121 PFTYGLWTPNVPVMTVGOEVRKTEGGQVSNLTGTGRLFMSSSAAMVGOPEDESKL 180
OY 121 PFTYGLWTPNVPVMTVGOEVRKTEGGQVSNLTGTGRLFMSSSAAMVGOPEDESKL 180
DB 121 PFTYGLWTPNVPVMTVGOEVRKTEGGQVSNLTGTGRLFMSSSAAMVGOPEDESKL 180
OY 181 PLFQFINWVKYKYPGCGEGSGDFTLDWTFDFDGSRMCKGWTDDGNRVDTLDKNI 240
DB 181 PLFQFINWVKYKYPGCGEGSGDFTLDWTFDFDGSRMCKGWTDDGNRVDTLDKNI 240
OY 181 PLFQFINWVKYKYPGCGEGSGDFTLDWTFDFDGSRMCKGWTDDGNRVDTLDKNI 240
DB 181 PLFQFINWVKYKYPGCGEGSGDFTLDWTFDFDGSRMCKGWTDDGNRVDTLDKNI 240
OY 241 YSRDGLMLALTRKQESFNGVPRDDEPAPOSSSSAPASSSVPASSSSAFV 300
DB 241 YSRDGLMLALTRKQESFNGVPRDDEPAPOSSSSAPASSSVPASSSSAFV 300
OY 301 PSSSSATNAIHGMRTTAVAKEHNLVNAKAKVNPNGHKRYRVNFPH 349
DB 301 PSSSSATNAIHGMRTTAVAKEHNLVNAKAKVNPNGHKRYRVNFPH 349
RESULT 2
H84053
endo-beta-1,3-1,4 glucanase (licheninase) bglIS [imported] - Bacillus halodurans (stra
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000

C:Accession: H84053
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasakl, R.; Masui, N.; Fujl, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20263314
 A:Accession: H84053
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-851 <STO>
 A:Cross-references: GB:AP001518; GB:BA000004; NID:910175792; PIDN:BA06951.1; GSPDB:GN0C
 A:Experimental source: strain C-125
 C:Genetics: dgls

Query Match 14.7% Score 268.5; DB 2; Length 851;
 Best Local Similarity 29.4%; Pred. No. 1.9e-11;
 Matches 83; Conservative 32; Mismatches 112; Indels 55; Gaps 12;

21 LTTNVAKDFGAEIYLYEYQKFEARMKMAAGSTVSMFLYONGSEIADGRPWVEV 80
 93 LENEQAPRYAGELRTNDFYQYGLFEVSMKPAVEGVTSFFTY-TGEMWDGDDPMDI 151
 81 DIEVLKNGPSQSNITTKAGAKTSEKHAIVSPADQAFHTYGLEWTPVYVMTVDCQ 140
 152 DIEFLGDTTRIQFNFTNGVG--NEFYDLGFDASESFNTYAFEMREDSITVYNGE 208
 141 EVRTBEGOVSNLTGT-OGLEFNLMSSSA-AWGOFDKSLPLFOFLNWKVYKYP-- 196
 209 AVHTA---TENIPQTKIMNLMVGVDEMGVDEGDTPLHAYEWY--RYTPLE 261
 197 --GQEGSGPFTLDWT-----DNFTDFDSRMC-KGDMTFDG 230
 262 ALDEESGDDEEPEVEEPEADEEVSVPISAIYEFNFNFEDINSIANGWI-NG 320
 231 NRDLT--DKNIYSRDCMLIALTRKQESFNGVPPDDEPA 270
 321 QMFNATWYDSQITFSNGIMRPAIDKE-----DDEEA 351

RESULT 3
 A48378
 Licheninase (EC 3.2.1.73) - *Bacillus brevis*
 N:Alternate names: lichenase
 C:Species: *Bacillus brevis*
 C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
 R:Low, M.E.; Reid, S.J.; Watson, T.G.
 Appl. Microbiol. Biotechnol. 38, 507-513, 1993
 A:Title: Characterization, cloning and sequencing of a thermostable endo-(1,3-1,4) beta-
 A:Reference number: A48378; MUID:93159752
 A:Accession: A48378
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-252 <JOU>
 A:Note: sequence extracted from NCBI backbone (NCBI:124672, NCBI:124673)
 C:Superfamily: licheninase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 14.6% Score 265.5; DB 2; Length 252;
 Best Local Similarity 35.8%; Pred. No. 6.7e-12;
 Matches 62; Conservative 24; Mismatches 74; Indels 13; Gaps 6;
 26 SAUDFSAELYTEYQKFEARMKMAAGSTVSMFLYONGSEIADGRPWVEV 85
 90 SANRYKAGELRTNDFYQYGLFEVSMKPAVEGVTSFFTY-TGEMWDGDDPMDI 148
 86 GKNPGFSQSNITTKAGAKTSEKHAIVSPADQAFHTYGLEWTPVYVMTVDCQ 145
 149 GKPTTRQFNFTNGVG--NEFYDLGFDASESFNTYAFEMREDSITVYNGEAVHTA 205
 146 EGGOVSNLTGT-OGLEFNLMSSSA-AWGOFDKSLPLFOFLNWKVYKYP 196

Db 206 ----TENIPQTKIMNLMVGVDEMGVDEGDTPLHAYEWY---RTPE 251

RESULT 4
 LXRBS
 Licheninase (EC 3.2.1.73) precursor - *Bacillus subtilis*
 N:Alternate names: 1,3-1,4-beta-D-glucan 4-gluconohydrolase; beta-glucanase; lichenas
 N:Contents: licheninase E-1; licheninase E-2
 C:Species: *Bacillus subtilis*
 C:Date: 30-Jun-1988 #sequence_revision 13-Mar-1998 #text_change 16-Jun-2000
 A:Accession: I40370; B69594; A22914; A90026; A90027; J00110
 R:Molf, M.; Geczi, A.; Simon, O.; Boriss, R.
 Microbiology 141, 281-290, 1995
 A:Title: Genes encoding xylan and beta-glucan hydrolysing enzymes in *Bacillus subtilis*
 A:Reference number: I40370; MUID:95219081
 A:Accession: I40370
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-242 <RES>
 A:Cross-references: EMBL:246862; NID:9599673; PIDN:CAA86922.1; PID:9599674
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritzt, C.; Fujita, M.; Fujita, Y.; Fume, S.; Gallizi, A.; Gal
 lech, J.; Harwood, C.R.; Henault, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowski, A.; Se
 kench, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumslein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: B69594
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-242 <RUN>
 A:Cross-references: GB:299124; GB:AL009126; NID:92636442; PIDN:CAB15943.1; PID:926364
 A:Experimental source: strain 168
 R:Murphy, N.; McConnell, D.J.; Cantwell, B.A.
 Nucleic Acids Res. 12, 5355-5367, 1984
 A:Title: The DNA sequence of the gene and genetic control sites for the excreted B. s
 A:Reference number: A93526; MUID:84272222
 A:Accession: A22914
 A:Molecule type: DNA
 A:Residues: 1-203, 'L', 205-242 <MUR>
 A:Cross-references: EMBL:X00754; NID:939818; PIDN:CAA25328.1; PID:9685236
 A:Experimental source: strain C120
 A:Note: the authors translated the codon CAA for residue 29 as Lys and CCA for residu
 R:Tezuka, H.; Yunk, T.; Yabuchi, S.
 Agric. Biol. Chem. 53, 2335-2339, 1989
 A:Title: Construction of a beta-glucanase hyperproducing *Bacillus subtilis* using the
 A:Reference number: A90026
 A:Accession: A90026
 A:Molecule type: DNA
 A:Residues: 1-23, 'S', 25-82, 'S', 84-242 <TEZ>
 A:Cross-references: DBJ:D00518; NID:9216243; PIDN:BA00405.1; PID:9216244
 A:Experimental source: strain Y-25, clone PLE100
 R:Yunk, T.; Tezuka, H.; Yabuchi, S.
 Agric. Biol. Chem. 53, 2341-2346, 1989
 A:Title: Purification and some properties of two enzymes from a beta-glucanase hyperp
 A:Reference number: A90027
 A:Contents: annotation
 A:Note: source was hyperproducing strain HU-25 with gene from strain Y-25
 C:Genetics: dgls
 A:Gene: dgls
 C:Function: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenin an

R:Schlammig, S.; Schwarz, W.H.; Staudenbauer, W.L.
Eur. J. Biochem. 204, 13-19, 1992
A:Title: Structure of the Clostridium thermocellum gene licB and the encoded beta-1,3-1,
l cellulases.
A:Reference number: S23498; MUID:92155194

A:Accession: S23498
A:Molecule type: DNA
A:Residues: 1-334 <SCH>
A:Cross-references: EMBL:X63355; NID:g40697; PIDN:CAA44959.1; PID:g40698
C:Genetics:

A:Gene: licB
C:Superfamily: licheninase licB; Clostridium cellulase repeat homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-334/Product: licheninase #status predicted <MNT>
F:273-296/Domain: Clostridium cellulase repeat homology <CCRI>
F:308-331/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 13.0%; Score 237; DB 1; Length 334;
Best Local Similarity 30.9%; Pred. No. 1e-09;

Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;

QY 28 KDFSGALYLYLEVO-----YGFEMAKMAAAGTVSSMFLYONGSEIADGRPWVEVD 81
DB 82 REYGSYPYKSGEYRTKSFEGYGEYVEMAKAKNVGAISSFFTYGPS---DNMPDEID 138
QY 82 IEVIGKRPDSFOSNITIGKGAQKTSKHNHVAAPADAFITYGLEMTVPRNYRTVTCOE 141
DB 139 IEFLGKDTTKVQFWMYKNGVG---NEYLHNLGFDASODFTTYGFEMPPDIDIFYDGKK 195
QY 142 VRKTEGQVSNLTGTGQ-LRFNLMSSSA-AWQGFDESKLPLOFIMVYV----- 192
DB 196 VYR-----GTRNIPYTPCKIMNMLPVGIGVDMLGRYD-GRTPLOAEYEVYVYNGVPQD 250
QY 193 -----KTPP-----GQGE-GGSDFT-----LDWTDFEFDGSRMGKG 224
DB 251 NPPTPTIATSTPNNPLPLKGDVNGDGHVNSDSYLFKRYLLRVIDRFPVGDQS---VA 307
QY 225 DMFPGNRVDLTOKNIYSR 243
DB 308 DVNRDG-RIDSTDLMLKR 325

RESULT 9

licB (EC 3.2.1.73) - Bacillus licheniformis
N:Alternate names: endo-beta-1,3-1,4-D-glucanase; lichenase
C:Species: Bacillus licheniformis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S15388
R:Roberts, J.; Perez-Pons, J.A.; Querol, E.
Eur. J. Biochem. 197, 337-343, 1991
A:Title: Molecular cloning, expression and nucleotide sequence of the endo-beta-1,3-1,4-
A:Reference number: S15388; MUID:91224124
A:Accession: S15388
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-243 <LIO>
A:Cross-references: EMBL:X57279; NID:g39558; PIDN:CAA0547.1; PID:g39559
C:Superfamily: licheninase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 12.8%; Score 233.5; DB 1; Length 243;
Best Local Similarity 33.7%; Pred. No. 1.2e-09;

Matches 58; Conservative 26; Mismatches 77; Indels 11; Gaps 6;

QY 23 TNSAKDFSGALYLYLEVOYGFEMAKMAAAGTVSSMFLYONGSEIADGRPWVEVD 82
DB 81 TSPSYNKFDCGKNSVPTTYGILEVNMKPAKNGVSSFFTYGPT---DGRPWDEID 137
QY 83 EVLGNKPGSFOSNITIGKGAQKTSKHNHVAAPADAFITYGLEMTVPRNYRTVTCOE 142

DB 138 EFLGKDTTKVQFNYNTNGVG---NHEKIVNLGFPAAANSYHTYARDQPNISIKRWVDGO-L 193
QY 143 KRTGCGOVSNLTGTGQLRFNLMSSSA-AWVGQDESKLPLOFIMVYVYK 193
DB 194 KHTATQIPQTPGK--IMMNLNMGAGVDEWLGSYN-GVTPLSRLSHWRYTK 242

RESULT 10

S64507
probable membrane protein YGR189c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G7553
C:Species: Saccharomyces cerevisiae
C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 29-Oct-1999
C:Accession: S64507
R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64499
A:Molecule type: DNA
A:Accession: S64507
A:Residues: 1-507 <ARR>
A:Cross-references: EMBL:Z72974; NID:g1323335; PID:e243566; PID:g1323336; GSPDB:GN000
A:Experimental source: Strain S288C
C:Genetics:
A:Gene: MIPS:YGR189c
A:Map position: 7R
C:Keywords: transmembrane protein
F:6-22/Domain: transmembrane #status predicted <TM>

Query Match 12.6%; Score 230; DB 2; Length 507;
Best Local Similarity 26.9%; Pred. No. 5.6e-09;

Matches 87; Conservative 57; Mismatches 134; Indels 46; Gaps 13;

QY 6 TAVKALAVAAAAALTTNYSKDFSGALYLYLEVOYGFEMAKMAAAGTVSSMFLY 65
DB 70 TDLKHAGEIKYSGDGLSMTL-AKRYDNPISLKSNEYIWKGLVILKAACTGIVSSFLQ 128
QY 66 QNGSEIADGRPWVEVDEVLGKNGPSFOSNITIGKGAQKTSKHNHVAAPADAFITYG 125
DB 129 SDLDL-----EIDLEWVGDTQFOSNFSKGTOTTYDRGEFGVDPTR-FHNTY 179
QY 126 LEWTPNYRTVTDQEVKRTGCGOVSNLTGTGQ-----LRFNLMSS--ESNA-----W 171
DB 180 LDNAMDKTWYLDGESVR-----VLSNTSSEGVPOSPMYLMMGYMGDPDMMAGTIEM 233
QY 172 V-QGFDESKLPLOFIMVYVYKTPGQEGSGDFTLDWTDFEFDGSRMGKDWTFDG 230
DB 234 AGGETNYNDAPFTMYIEKIVTVDTGKRYTYGQSGS-ESIEADGSIYGRYDQAE- 291
QY 231 NRVDLTGKNIYSRDMILALTRKG--QESFNGOVPRDDEPAPSSSSAPASS--SSVPA 286
DB 292 -----DFAVLANGSISSSTSSSTSYSSASSTVSSSVSSSTVSSSVSSSVTS 345
QY 287 SSSSVPASSSSAFVPPSSSATNA 310
DB 346 SSSSVSSSSSTS---PSSSTATSS 366

RESULT 11

S11927
licheninase (EC 3.2.1.73) precursor [validated] - Bacillus macerans
N:Alternate names: endo-beta-1,3-1,4-glucanase; lichenase
C:Species: Bacillus macerans
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Sep-2000
C:Accession: S11927
R:Boriss, R.; Bueltner, K.; Maentzsaee, P.
Mol. Gen. Genet. 222, 278-283, 1990
A:Title: Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans: homologie
A:Reference number: S11927; MUID:91109712
A:Accession: S11927
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-237 <BOR>
A:Cross-references: EMBL:X55959; NID:g296715; PIDN:CA39426.1; PID:g296716
R:Hahn, M.; Heinemann, U.
submitted to the Brookhaven Protein Data Bank, December 1994
A:Reference number: A67074; PDB:LMC
A:Contents: annotation; X-ray crystallography, 2.3 angstroms, residues 26-37, 'P', 39-237
A:Note: recombinant form with residues 26-41 from *Bacillus amyloliquefaciens* expressed in
R:Kistler, T.; Simon, O.; Borriess, R.; Heinemann, U.
Proc. Natl. Acad. Sci. U.S.A. 90, 5287-5291, 1993
A:Title: Molecular and active-site structure of a *Bacillus* 1,3-1,4-beta-glucanase.
A:Reference number: A47562; MIMD:93281743
A:Contents: annotation; X-ray crystallography, 2.3 angstroms
A:Note: recombinant form with residues 26-41 from *Bacillus amyloliquefaciens* expressed in
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenin and
C:Superfamily: licheninase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-25/Domain: signal sequence #status predicted <PRO>
F:26-237/Product: licheninase #status experimental <MAT>
F:55-84/Disulfide bonds: #status experimental
F:120/Active site: Glu #status predicted

Query Match	12.5%	Score 228;	DB 1;	Length 237;
Best Local Similarity	34.3%	Pred. No. 2.9e-09;		
Matches 60;	Conservative 25;	Mismatches 74;	Indels 16;	Gaps 8;

OY 23 TNSAKDFSCAELTYLEEVOCKFEARKKMAAAAGTVSSMFLYONGSEIADGRPWYEVDI 82
 | : : | : : | : : | : : | : :
Db 75 TTSAVNNFPCAERYRNIGYGLEYEVSMMPAKNNGIVSSFETLYGP--AHGTQWMEIDI 131

```
QY      83 EVLGKNPGSQSNILTGKAGAKTSEKHHAVSADQAHTHYGLEMTPNRYRWTVGQEV 142
          +   +   +   +   +   +   +   +   +   +   +   +   +   +   +   +
Db     132 EFLGKDITKVQFNYYTNGVGHH--EKVISLGEADASKGFHTYAEPDWPQGIYIKWYVDG-VL 187
```

```

QY      143 RTEGGGVSNLTGQG-LRFNLMSSESA-AWVGQDESKLPLEQFINNWKYKKT 195
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      188 KHT--ATANIPSPGKIMMNIWNGTGVDDWLGSYNGAN-PLAELYDWV---KYT 235

```

RESULT 12
S19012
l-phenylalanine (EC 3.2.1.73) precursor - Bacillus notumura

C:Accession: S19012
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 18-Jun-1999
C:Species: Bacillus polymyxa
N:Alternate names: endo-beta-L,3-1,4-glucanase; lichenase

A:Reference number:SI9011; MUID:92041687

A:Accession: 5525012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-238 <GOS>

C:Function: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenin and
C:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenin and
C:Superfamily: licheninase

F;129/Active site: Glu #status predicted
F;56-85/Disulfide bonds: #status predicted
F;27-237/Product: licheninase #status predicted <MAT>
F;129/Domain: signal sequence #status predicted

Query Match	12.58;	Score 227.5;	DB 1,	Length 238;
Best Local Similarity	34.38;	Pred. No. 3.2e-09;		
Matches 61; Conservative	26;	Mismatches 74;	Indels 17;	Gaps 9;

Qy 20 ALTTNVSAKDFSGAELTYTLEEVOYGFKEAPRMKMAASGTIVSSMFLYONGSEIADGRPAVE 79
:: : : : : : : : : : : : : : : : : :
Db 74 SLTSVAANK-PDCCGEYRSTNNNGYGLYEYSMKPAAKTGTVSSFFYYTGPSH---GTOWDE 129

[illegible]

RESULT 13

endo-1,3(4)-beta-glucanase (EC 3.2.1.6) precursor - Clostridium thermocellum
N/Alternate names: laminarinase
C/Species: Clostridium thermocellum

C:Accession:JS0611; S18726
R:Zverlov, V.V.; Laptev, D.A.; Tishkov, V.I.; Velikodvorskaja, G.A.
Biochem. Biophys. Res. Commun. 181, 507-512, 1991

A;Reference number: JS0611; MUID:92095946
A;Accession: JS0611
A;Molecule type: DNA
A;Postfix: 1 243 -29815

A; Cross-references: EMBL: X58392
R; Zverlov, V.V.; Velikodvorskaya, G.A
Biotechnol. Lett. 12, 811-816, 1990
A; Title: Cloning the Clostridium thermophilum

A:Reference number: S18726
A:Accession: S18726
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: RNA

A:Residues: 1-242 <ZVE2>
A:Cross-references: EMBL:X58392
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1991
C:Genetics:

C;Superfamily: licheninase
A;Gene: lam1
C;Keywords: glycosidase, hydrolase, polysaccharide degradation
E:1-27/Domaln:signal sequence #status predicted <sig>

```

E: 26-24Z/PRODUCT: enao-1,3(4)-beta-glucanase *status predicted <END>
Query Match      12.3%:  Score 224.5:  DB 2:  Length 242:

```

OY	14 VAAAAAATTNVS-----YGFPEARMKMAASGTV	59
	AKDFSGAEIYLTEEVQ-----	
	Best保守率 50.0%	77
	Mismatches 62; Conservative 25; Indels 27; Gaps 8	

Db 58 VSTVLEAFETDISNGKMLITLIDREYGGSPYKSGEYRTSSFFGGYGEVBRMAKAKNGV 117
QY 60 SSMELYQNGSEIADGRPRWEVDIEVLCKNPGSPFSQNIITTKAGAQKTEKHHAVSPADQ 119

Db 118 SFEFTYTGPS--DNNPWDEIDIEFLGKDTKYQFNWIKNGVGG---NEYLHNLGFDASQ 177

Qy 120 AFHTTGLEWTPNYVRMTVDGQEVKRTGEGQVSNLTGTQG-LRFNLSSESA-AMVGQFDE 177

Db	172	DEHTYGFEMRPDIIDYVDGKKVYR-----GTRNIPVTPGKIMMNLMPGIGVDEWLGRID-22
QY	178	SKLPL182

Db 227 GRPL 231

g84053
hybrid-endo-beta-1,3-1,4 glucanase BH3231 [imported] - *Bacillus halodurans* (strain C-1)
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence revision 01-Dec-2000 #text change 31-Dec-2000

R. Takami, H.; Takasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Nucleic Acids Res. 28, 4317-4331, 2000

Article: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a

A:Reference number: A83650; MUID:20263314
 A:Accession: G84053
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-302 <STO>
 A:Cross-references: GB:AP001518; GB:BA000004; NID:910175792; PIDN:BAB06950.1; GSPDB:GNOC
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3231

Query Match 10.4% Score 190.5; DB 2; Length 302;
 Best Local Similarity 27.8%; Pred. No. 1.9e-06;
 Matches 62; Conservative 34; Mismatches 98; Indels 29; Gaps 9;
 QY 29 DFGSAELTYLEEVYGFKEARMKMAASGVSSMELYNONGSELDGPRWEVDIEVGN 88
 16 NYCEAEVSTHNFYGYGLXEVSMKSNVSVISFFTYTGS--YNGAPMDEIDIEFGND 73
 QY 89 PGSPQSNITGKGAQKTSKHHAVSPADQAFHTYGLLEMTPNVRYTVGQEVKTEGG 148
 DB 74 TTKVQFNYTYNGRG--NELYDLGFDANSEFTYAFDMQEHYINMYVNGELVATATEN 130
 QY 149 QVSNLTGTGRLRNLMSF-----SAWVGQFDESKLPLEQFINWVYKYYTP---GCG 199
 DB 131 IPSN---PSKMMNMIMNTYVIDEWAGAYWQONANAS-----YEWV---RYTPNERSQT 178
 QY 200 ECGSDFTLDWTNDFPDGSRM--GKGDWTFEGNRYDLTDKNI 240
 DB 179 PIASDPQLHACEYSDANGVTSWCGVGSF--YPSMWIKFDVNL 220

RESULT 15

S30839
 UTR2 protein - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YEL040w
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 06-Feb-1998
 C:Accession: S30839; S50504; S38545
 R:Multigan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wel, Y.; Taylor, submitted to the EMBL Data Library, February 1993
 A:Reference number: S30812
 A:Accession: S30839
 A:Molecule type: DNA
 A:Residues: 1-467 <MUL>
 A:Cross-references: GB:U18779; EMBL:L10830; NID:9603625; PID:9603639
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.
 A:Reference number: S50491
 A:Accession: S50504
 A:Molecule type: DNA
 A:Residues: 1-467 <DIE>
 A:Cross-references: EMBL:U18779; NID:9603625; PID:9603639; MIPS:YEL040w
 R:Melnick, L.; Sherman, F.
 J. Mol. Biol. 233, 372-388, 1993
 A:Title: The gene clusters ARC and COR on chromosomes 5 and 10, respectively, of Sacchar
 A:Reference number: S38543; MUID:94016558
 A:Accession: S38545
 A:Molecule type: DNA
 A:Residues: 121-129, 'V', 131-290, 'R', 292-353, 'C', 355-467 <MEL>
 A:Cross-references: EMBL:S66130; NID:9430829; PID:9430830
 C:Genetics:
 A:Gene: SGD:UTR2
 A:Cross-references: SGD:S0000766; MIPS:YEL040w
 A:Map position: 5L

123 KNSGFTLSTRAWVKGVARIKTSHLAGVYTGFIYSGAGD-----ELDYEFVGA 174
 QY 88 NPGSPQSNITGKGAQKTSKHHAVSPADQAFHTYGLLEMTPNVRYTVGQEVKTEGG 140
 DB 175 DLETAQTFYVESYLYNTNSANISTD-----TFENVHRYELDMHEDVYTWSIDGV 225
 QY 141 EVRKTEGGQVSNLTGQGLRF-----NLW-----SSESA-----AWVG-----OPDESKLP 181
 DB 226 VGRFLYKNETYNAT-PQKYQYPTPSKVDISIMPGCNSTAPGTIMSGGEIMWDAISDIS 284
 QY 182 ----LFOFINWVYKYYKTPGQGBGSDFTLDWTFNEDTFDGSRWGKGDWTFDGNRYDLTD 237
 DB 285 NPGYVAIVNEVNTCYDP-----PSDTKKKNGTSAYVYTSSE-----FLAKDIAITD 332
 QY 238 KNIY--SRDGLLIALLRKQESFNGVPRDDEPAPQSSSAPASSSVPAS-----287
 DB 333 DEVMQSDDEG-----SLDPRKGTAT--TSYQKSSSTATSSSKTSSDHSSTKK 380
 QY 288 ---SSVPSASSSAFVPPSSSATMAIGHMRTTPAVA 321
 DB 381 SSKTSSSTASSSSSSSSSSSSSSSTATFKNGDKVYSSVS 417

Search completed: August 23, 2001, 11:15:08
 Job time: 95 sec

Query Match 10.4% Score 189.5; DB 2; Length 467;
 Best Local Similarity 24.6%; Pred. No. 3.9e-06;
 Matches 83; Conservative 49; Mismatches 120; Indels 85; Gaps 15;
 QY 28 KDFSGAELTYLEEVYGFKEARMKMAASGVSSMELYNONGSELDGPRWEVDIEVLGK 87

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:16:31 ; Search time 19.66 Seconds

(without alignments)
608.096 Million cell updates/sec

Title: US-09-654-652a-3
Perfect score: 1824
Sequence: 1 MNKKTKAVKSAALAVAAAAA.....AKGAKVNPNGHKRYVNEEH 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1824	100.0	349	1 GUB_FIBSU	P17989 fibrobacter
2	266.5	14.6	259	1 GUB_BACR	P37073 bacillus br
3	248	13.6	242	1 GUB_BACSU	P04957 bacillus su
4	242.5	13.3	239	1 GUB_BACAM	P07980 bacillus am
5	242.5	13.3	802	1 XYND_RUMFL	O53317 ruminococcu
6	237	13.0	334	1 GUB_CLOTH	P29716 clostridium
7	233.5	12.8	243	1 GUB_BACLI	P27051 bacillus li
8	230	12.6	507	1 YG46_YEAST	P33301 saccharomyc
9	228	12.5	237	1 GUB_PAEWA	P23904 paenibacill
10	227.5	12.5	238	1 GUB_PAEPO	P45797 paenibacill
11	197	10.8	269	1 EXOK_RHIME	P33693 rhizobium m
12	189.5	10.4	347	1 UTR2_YEAST	P32623 saccharomyc
13	154	8.4	247	1 MER2_ARATH	P24806 arabidopsis
14	133.5	7.3	286	1 GUB_RHOMR	P45780 rhodothermu
15	128	7.0	283	1 BRUI_SOYBN	P35694 glycinine max
16	119	6.5	465	1 EXSH_RHIME	O33680 rhizobium m
17	117	6.4	397	1 CGKA_ALTCA	P43498 alteromonas
18	115	6.3	682	1 EL3B_BACCI	P23903 bacillus ci
19	108.5	5.9	969	1 SERA_PLAFG	P13823 plasmodium
20	107	5.9	484	1 YODO_YEAST	O08193 saccharomyc
21	105.5	5.8	719	1 PBPA_STRPN	O04707 streptococc
22	104	5.7	962	1 GUNA_PSEFL	P10476 pseudomonas
23	103.5	5.7	1580	1 ACC8_HUMAN	O09438 homo sapien
24	102.5	5.6	879	1 GUN1_CLOTH	O02934 clostridium
25	102	5.6	465	1 EGIC_RHIME	O93342 rhizobium m
26	102	5.6	650	1 SC02_NEUCR	O01277 neurospora
27	101	5.5	691	1 VIVI_MAZE	P26307 zea mays (m
28	100.5	5.5	1036	1 NIT2_NEUCR	P19212 neurospora
29	100	5.5	748	1 GUNC_PSEFL	P27033 pseudomonas
30	99	5.4	560	1 YDEN_ECOLI	P77318 escherichia
31	98.5	5.4	658	1 GUN3_FIBSU	P14250 fibrobacter
32	98	5.4	1845	1 Z236_HUMAN	O9136 homo sapien
33	97.5	5.3	1385	1 C5AA_BACUD	O45760 bacillus th

34	97.5	5.3	1580	1 ACC8_RAT	O09429 ratius norv
35	97.5	5.3	1581	1 ACC8_CRICR	O09427 cricetus cr
36	96.5	5.3	526	1 CAP_YEAST	P17555 saccharomyc
37	96	5.3	371	1 LIG2_PHACH	P49012 phanerocha
38	95.5	5.2	872	1 IP3L_RAT	P42335 ratius norv
39	95.5	5.2	1289	1 C5AB_BACUD	O45753 bacillus th
40	95	5.2	400	1 GUN5_BACAG	O85465 bacillus ag
41	94.5	5.2	451	1 MYBH_DICDI	P34127 dictyostell
42	94.5	5.2	586	1 SYN2_RAT	O63537 ratius norv
43	94.5	5.2	928	1 KKL1_YEAST	P36003 saccharomyc
44	94	5.2	409	1 NCAP_IBVK	P12648 avian infec
45	93.5	5.1	619	1 VHP1_CABEL	Q10038 caenorhabdi

ALIGNMENTS

RESULT	ID	Query Match	Length	ID	Description
1	GUB_FIBSU	100.0	349	1 GUB_FIBSU	P17989 fibrobacter
AC	P17989	100.0	349	1 GUB_FIBSU	P17989 fibrobacter
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)				
DE	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).				
OS	Fibrobacter succinogenes (Bacteroides succinogenes).				
OC	Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;				
OC	Fibrobacter.				
OX	NCBI_TaxID=833;				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.				
RC	STRAIN-ISOLATE 585;				
RK	MEDLINE=90299807; PubMed=2193918;				
RA	Teather R.M., Erfle J.D.;				
RT	"DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";				
RL	J. Bacteriol. 172:3837-3841(1990).				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.				
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation.				
CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; M33676; AAA24896.1; -				
DR	PIR; A44507; A44507.				
DR	HSSP; P23904; 1AJK.				
DR	InterPro; IPR000757; -				
DR	Pfam; PF00722; Glyco_Hydro.16; 1.				
DR	PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.				
FW	Hydrolase; Glycosidase; Signal; Repeat.				
FT	SIGNAL	1	27		
FT	CHAIN	28	349		
FT	ACT_SITE	79	79		
FT	ACT_SITE	83	83		
FT	DOMAIN	271	307		
FT	REPEAT	271	277		
FT	REPEAT	278	284		
FT	REPEAT	285	291		
FT	REPEAT	292	298		
FT	REPEAT	301	307		
SO	SEQUENCE	349 AA;	37737 MM;		

Query Match 100.0%; Score 1824; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 8.8e-122;

Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKTAVKALVAAAAAALTTNSKDSGAEVLYEEVQKGFARKMAASTVS 60
 1 MNKKTAVKALVAAAAAALTTNSKDSGAEVLYEEVQKGFARKMAASTVS 60
 Db 1 MNKKTAVKALVAAAAAALTTNSKDSGAEVLYEEVQKGFARKMAASTVS 60

QY 61 SMFLYONGSEIADGRPWVEVDIEVLGNKPSQSNITTKAGAKTSEKHHAVSPADQA 120
 61 SMFLYONGSEIADGRPWVEVDIEVLGNKPSQSNITTKAGAKTSEKHHAVSPADQA 120
 Db 61 SMFLYONGSEIADGRPWVEVDIEVLGNKPSQSNITTKAGAKTSEKHHAVSPADQA 120

QY 121 FHNYGLEMPNRYRWYVDGGEVARKTEGGVSNLTGDLFNLMSSSAAWGOFPDESKL 180
 121 FHNYGLEMPNRYRWYVDGGEVARKTEGGVSNLTGDLFNLMSSSAAWGOFPDESKL 180
 Db 121 FHNYGLEMPNRYRWYVDGGEVARKTEGGVSNLTGDLFNLMSSSAAWGOFPDESKL 180

QY 181 PLTQFIMNVVYRYTTPQGGSGSDFTLDMTNDPTFDGSRWKGMDTFDGNRYDLTDKNI 240
 181 PLTQFIMNVVYRYTTPQGGSGSDFTLDMTNDPTFDGSRWKGMDTFDGNRYDLTDKNI 240
 Db 181 PLTQFIMNVVYRYTTPQGGSGSDFTLDMTNDPTFDGSRWKGMDTFDGNRYDLTDKNI 240

QY 241 YSRDGMILIALTRKGOSEFNGQVPRDEPAPQSSSSAPASSSSVPASSSSAFV 300
 241 YSRDGMILIALTRKGOSEFNGQVPRDEPAPQSSSSAPASSSSVPASSSSAFV 300
 Db 241 YSRDGMILIALTRKGOSEFNGQVPRDEPAPQSSSSAPASSSSVPASSSSAFV 300

QY 301 PSSSSATNAINHGKRTTPAVAKERHNLVNAKAKVNPNGKRRVNEFH 349
 301 PSSSSATNAINHGKRTTPAVAKERHNLVNAKAKVNPNGKRRVNEFH 349
 Db 301 PSSSSATNAINHGKRTTPAVAKERHNLVNAKAKVNPNGKRRVNEFH 349

RESULT 2
 GUB_BACBR STANDARD; PRT: 259 AA.

AC P37073;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE 01-JUN-1994 (Rel. 29, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 GN (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 OS BGLB.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Brevibacillus.
 OX NCBI_TaxID=1393;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93159752; PubMed-7763386;
 RA Louw M.E., Reid S.J., Watson T.G.;
 RT "Characterization, cloning and sequencing of a thermostable
 RT endo-(1,3-1,4) beta-glucanase-encoding gene from an alkalophilic
 RT Bacillus brevis";
 RL Appl. Microbiol. Biotechnol. 38:507-513(1993).
 CC -1- FUNCTION: HYDROLYSES B-GLUCANS CONTAINING MIXED BETA-1,3 AND
 CC BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; ITS OPTIMAL
 CC TEMPERATURE IS NEAR 65-70 DEGREES CELSIUS.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; M84339; AAA22265.1; -
 DR HSSP; P23904; IAJK.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro.16; 1.
 DR PRINTS: PR00737; GLYHRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 259 BETA-GLUCANASE.
 FT ACT SITE 142 142 NUCLEOPHILE (BY SIMILARITY).
 FT ACT SITE 146 146 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 259 AA; 29960 MM; A630C9F281FF5D13 CRC64;

Query Match 14.6%; Score 266.5; DB 1; Length 259;
 Best Local Similarity 36.0%; Pred. No. 3,9e-12;
 Matches 63; Conservative 24; Mismatches 75; Indels 13; Gaps 6;

QY 26 SAKDFSGAEVLYEEVQKGFARKMAASTVS SMFLYONGSEIADGRPWVEVDIEVL 85
 26 SAKDFSGAEVLYEEVQKGFARKMAASTVS SMFLYONGSEIADGRPWVEVDIEVL 85
 Db 26 SAKDFSGAEVLYEEVQKGFARKMAASTVS SMFLYONGSEIADGRPWVEVDIEVL 85

QY 90 SARNYKAGELRTNDFYHYGLFEVSMKPAKVEGTYSFFTY-TGEMDMDGDPMEIDIEFL 148
 90 SARNYKAGELRTNDFYHYGLFEVSMKPAKVEGTYSFFTY-TGEMDMDGDPMEIDIEFL 148
 Db 90 SARNYKAGELRTNDFYHYGLFEVSMKPAKVEGTYSFFTY-TGEMDMDGDPMEIDIEFL 148

QY 86 GRNPGSQSNITTKAGAKTSEKHHAVSPADQA FHTYGLTPTPYVRYVDGGEVARKT 145
 86 GRNPGSQSNITTKAGAKTSEKHHAVSPADQA FHTYGLTPTPYVRYVDGGEVARKT 145
 Db 86 GRNPGSQSNITTKAGAKTSEKHHAVSPADQA FHTYGLTPTPYVRYVDGGEVARKT 145

QY 149 GKDTRIQFYFTNGVGG--NEFYVDLGFDASESFNFAFEWREDSITVYVNGEAVHTA 205
 149 GKDTRIQFYFTNGVGG--NEFYVDLGFDASESFNFAFEWREDSITVYVNGEAVHTA 205
 Db 149 GKDTRIQFYFTNGVGG--NEFYVDLGFDASESFNFAFEWREDSITVYVNGEAVHTA 205

QY 146 EGGQVSNLTGT-GLRPNLWSSSA-AWGOFPDESKLPLTQFIMNVVYRYTTPQ 198
 146 EGGQVSNLTGT-GLRPNLWSSSA-AWGOFPDESKLPLTQFIMNVVYRYTTPQ 198
 Db 146 EGGQVSNLTGT-GLRPNLWSSSA-AWGOFPDESKLPLTQFIMNVVYRYTTPQ 198

QY 206 -----TENIPOTPOKIMNLMIPGVGDGWTGVFDGDNTPVYTYDWT---RTTPLQ 253
 206 -----TENIPOTPOKIMNLMIPGVGDGWTGVFDGDNTPVYTYDWT---RTTPLQ 253
 Db 206 -----TENIPOTPOKIMNLMIPGVGDGWTGVFDGDNTPVYTYDWT---RTTPLQ 253

RESULT 3
 GUB_BACSU STANDARD; PRT: 242 AA.

AC P04957;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLS OR BGL OR LICS OR N15B.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-C120;
 RX MEDLINE-84272222; PubMed-6087283;
 RA Murphy N., McConnell D.J., Cantwell B.A.;
 RT "The DNA sequence of the gene and genetic control sites for the
 RT excreted B. subtilis enzyme beta-glucanase";
 RL Nucleic Acids Res. 12:5355-5367(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-HL-25;
 RA Tezuka H., Yuuki T., Yabuuchi S.;
 RT "Construction of a beta-glucanase hyperproducing Bacillus subtilis
 RT using the cloned beta-glucanase gene and a multi-copy plasmid";
 RL Agric. Biol. Chem. 53:2335-2339(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-168 / BGSCIA1;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RL Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX STRAIN-168;
 RX MEDLINE-95219081; PubMed-7704256;
 RA Wolf M., Geczi A., Simon O., Boriss R.;
 RT "Genes encoding xylan and beta-glucan hydrolysing enzymes in Bacillus
 RT subtilis: characterization, mapping and construction of strains
 RT deficient in lichenase, cellulase and xylanase";
 RL Microbiology 141:281-290(1995).
 RN [5]
 RP SEQUENCE OF 1-6 FROM N.A.
 RX STRAIN-BR151;
 RX MEDLINE-96178961; PubMed-8606172;
 RA Schmetz K., Stuelke J., Getz S., Krueger S., Kriegl M., Hecker M.,
 RA Rak B.;

```

RT "Llct, a Bacillus subtilis transcriptional antiterminal protein of
RT the BglG family."
RT J. Bacteriol. 178:1971-1979(1996).
RN [6]
RP SEQUENCE OF 29-63.
RA Yuki T., Tezuka H., Yabuuchi S.;
RT "Purification and some properties of two enzymes from a beta-glucanase
RT hyperproducing strain, Bacillus subtilis HL-25."
RL Agric. Biol. Chem. 53:2341-2346(1989).
CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASES OF GERMINATING BARLEY.
CC -I- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X00754; CAA25328.1; -
DR EMBL, D00518; BAA00405.1; -
DR EMBL, D83026; BAA11697.1; -
DR EMBL, Z46862; CAA86922.1; -
DR EMBL, Z28340; CAA82195.1; -
DR EMBL, Z59124; CAA15943.1; -
DR PIR, A22914; LXBS.
DR PIR, J00110; J00110.
DR HSSP, P27051; 1GBG.
DR Subtilist; BG10476; bglS.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 28
FT CHAIN 29 242
FT MOD_RES 29 29
FT ACT_SITE 133 133
FT ACT_SITE 137 137
FT DISULFID 60 89
FT VARIANT 24 24
FT VARIANT 83 83
FT VARIANT 204 204
SQ SEQUENCE 242 AA; 27268 MW; 45958DEA70F22B29 CRC64;

Query Match 13.6%; Score 248; DB 1; Length 242;
Best Local Similarity 36.0%; Pred. No. 7.3e-11;
Matches 63; Conservative 25; Mismatches 75; Indels 12; Gaps 7;

OY 20 ALTVASAKDFSGALTYLIEVOYGFKFAKMAAASGVSMFLYKONSEIADGRPWVE 79
DB 78 ALTSAYVNF-FCGGERNSVYQYGLYERMRKPAKNTGVSFFYTGPT---DGPWDE 133
OY 80 VDIEVLGNKPGSFQSNILITGKAGAKTSEKHAHVSPADQAQHTYGLETPYVRYVTG 139
DB 134 IDIEFLGNDTTRVQFNYYTINGAG---NHEKIVDLGFDANAHHTAFAFDQPSIKRYV 190
OY 140 QEVKRTGGOVSNLTGTQGLRPNLWSSESA-AWVGQFDESKLPLFOFINWVYK 193
DB 191 Q-LKHTATNQTPTTPGK--IMMNLWNGTGVDMLGSYNGVN-PIYAHYDMVRYTK 241

RESULT 4
GUB_BACAM STANDARD; PRT; 239 AA.
AC GUB_BACAM P07980;

```

```

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN BGLA.
OS Bacillus amyloliquefaciens.
OC Bacillus; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BE 20/78;
RX Hofmeister J., Kurtz A., Borris R., Knowles J.;
RT "The beta-glucanase gene from Bacillus amyloliquefaciens shows
RT extensive homology with that of Bacillus subtilis."
RL Gene 49:177-187(1986).
CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -I- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASES OF GERMINATING BARLEY.
CC -I- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M15674; AAA87323.1; -
DR PIR, A29091; A29091.
DR HSSP, P27051; 1GBG.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 25
FT CHAIN 26 239
FT ACT_SITE 134 134
FT DISULFID 57 86
SQ SEQUENCE 239 AA; 26928 MW; A76A64268A7AAB0B CRC64;

Query Match 13.3%; Score 242.5; DB 1; Length 239;
Best Local Similarity 34.3%; Pred. No. 1.8e-10;
Matches 59; Conservative 27; Mismatches 75; Indels 11; Gaps 6;

OY 23 TVASAKDFSGALTYLIEVOYGFKFAKMAAASGVSMFLYKONSEIADGRPWVEVDI 82
DB 77 TSPSYNKFPCGGERNSVYQYGLYERMRKPAKNTGVSFFYTGPT---GTPDEIDI 133
OY 83 EVLGNKPGSFQSNILITGKAGAKTSEKHAHVSPADQAQHTYGLETPYVRYVTGGOV 142
DB 134 EFLGKDTTRVQFNYYTINGAG---NHEKPADLGFDANAHHTAFAFDQPSIKRYV 189
OY 143 RRTGGOVSNLTGTQGLRPNLWSSESA-AWVGQFDESKLPLFOFINWVYK 193
DB 190 KHTATTTQIPAAVAK--IMMNLWNGTGVDMLGSYNGVN-PIYAHYDMVRYK 238

RESULT 5
XYND_RUMFL
AC XYND_RUMFL STANDARD; PRT; 802 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE XYLANASE/BETA-GLUCANASE PRECURSOR (INCLUDES: ENDO-1,4-BETA-XYLANASE
DE (EC 3.2.1.8) (XYLANASE); (ENDO-BETA-1,3-1,4 GLUCANASE) (EC 3.2.1.73)

```

DR	EMBL:	S61204;	IAAKB.
DR	HSSP:	P23904;	AAK.
DR	InterPro:	IPR000757;	-
DR	InterPro:	IPR001137;	-
DR	Pfam:	PF02018;	CBD_6; 1.
DR	Pfam:	PF00457;	Glyco_hydro_11; 1.
DR	Pfam:	PF00722;	Glyco_hydro_16; 1.
DR	PRINTS:	PR00737;	GLHYDRLASE16.
DR	PRINTS:	PR00911;	GLHYDRLASE11.
DR	PROSITE:	PS00776;	GLYCOSYL_HYDROL_F11_1; 1.
DR	PROSITE:	PS00777;	GLYCOSYL_HYDROL_F12_1; 1.
DR	PROSITE:	PS01034;	GLYCOSYL_HYDROL_F16; 1.
KW	Xylan degradation:	Hydrolase;	Glycosidase; Signal;
KW	Multifunctional enzyme:		
FT	SIGNAL	1	31
FT	CHAIN	32	802
FT	DOMAIN	32	244
FT	DOMAIN	245	523
FT	DOMAIN	524	555
FT	DOMAIN	556	802
FT	ACT_SITE	124	124
FT	ACT_SITE	226	226
FT	ACT_SITE	684	684
FT	DOMAIN	524	529
FT	DOMAIN	532	543
FT	DOMAIN	546	553
QO	SEQUENCE	802 AA;	89091 MW; 2880A689647284AF CRC64;

[illegible]

RESULT 6			
ID	GUB_CLOTM	STANDARD:	PRT: 334 AA.
AC	P29716; P37074;		
DC	01-APR-1993 (Rel. 25, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	01-FEB-1996 (Rel. 33, Last annotation update)		
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)		
DE	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE) (LAMINARINASE).		
GN	LICB OR LAMI.		
OS	Clostridium thermocellum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_Taxid=1515;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-ATCC 27405 / DSM 1237;		
RC	MEDLINE=9215194; Pubmed=1740123;		
RA	Schliming S., Schwarz W.H., Staudenbauer W.L.;		
RT	"Structure of the Clostridium thermocellum gene licB and the encoded		
RT	beta-1,3-1,4-glucanase. A catalytic region homologous to Bacillus		
RT	lichenases joined to the reiterated domain of clostridial		
RL	cellulases.";		
RL	Eur. J. Biochem. 204:13-19(1992).		
RN	[2]		
RP	PRELIMINARY SEQUENCE FROM N.A.		

```

CC RA MEDLINE=92095946; PubMed=1755832;
CC RA Zverlov V.V., Laptev D.A., Tishkov V.I., Velikodvorskaja G.A.;
CC RT "Nucleotide sequence of the Clostridium thermocellum laminarinase
CC RL gene."
CC RL Biochem. Res. Commun. 181:507-512(1991).
CC [3]
CC REVISIONS.
CC RA Zverlov V.V.;
CC RL submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
CC [4]
CC CHARACTERIZATION.
CC RC STRAIN=F7;
CC RA Zverlov V.V., Velikodvorskaja G.A.;
CC RT "Cloning the clostridium thermocellum thermostable laminarinase gene
CC RL in Escherichia coli; the properties of the enzyme thus produced.";
CC RL Biotechnol. Lett. 12:811-816(1990).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- SUBUNIT: MAY FORM PART OF A MULTENZYME COMPLEX (CELLULOSE).
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).

```

DR EMBL; X63355; CAA44959.1; -
DR EMBL; X56392; CAA41281.1; -
DR PIR; S23498; S23498.
DR PIR; J50611; J50611.
DR PIR; S18726; S18726.
DR HSSP; P23904; IATK.

```

DR InterPro: IPR000757; -
DR InterPro: IPR002048; -
DR InterPro: IPR002105; -
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE: PS00448; CLOS_CELULOSOME_RPT; 2.
KW Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 334 BETA-GLUCANASE.
FT ACT_SITE 136 136 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 140 140 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 252 269 PRO/THR-RICH (LINKER).
FT DOMAIN 273 331 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 273 296 1.
FT REPEAT 308 331 2.
FT CONFLICT 304 334 QSVADVNRDGRIDSTDLMLKRLIRAIPL ->
FT PODCGRHDRVVDGSK (IN REF. 2 AND 3).
SQ SEQUENCE 334 AA; 37897 MW; 0837564E9726F281 CRC64;
Query Match 13.08; Score 237; DB 1; Length 334;
Best Local Similarity 30.9%; Pred. No. 6.4e-10;
Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;
QY 28 KDFSGAEIYLTLEEVQ-----YGFKEARMKMAASGVSMFLYONGSEIADGRPWVEVD 81
DB 82 REYGSGIYKKGKGYRTKSGFEFGIYEVKMAKANGIVISFEFTYGPS--DNNMDEID 138
QY 82 IEVLKKNFSFOSNITITKAGAKQKTEKHAHVSPADQAFHTYGLFWPNVYRWTVDSQE 141
DB 139 IEPFGKDTTKYQFMNMYKNGVGG--NEYLNLGPDASQDFHTYGEEMRPDIYDFVDSKK 195
QY 142 VRKIEGGGVSNLTGTQG-LRNLMSSESA-AWVGQFDESKLPLEFOFIWVKVY----- 192
DB 196 VYR---GTRNIPVTPGKIMNLMFGICVDMLGRYD-GRTPLQAEYEVYVYYPNGVPOD 250
QY 193 -----KTRP-----GQGE-GGSDPT-----LMDTMDPDPFDSGRMKKG 224
DB 251 NPPTPTTAPSTPTNPMLPLKGDVNGDSGVNSDSYLFKRYLLRVIDRPVGDQS---VA 307
QY 225 DWTEFGNREVDLTDKNIYSR 243
DB 308 DVNRDG-RIDSTDLMLKR 325
RESULT 7
GUB_BACLI STANDARD; PRT: 243 AA.
ID GUB_BACLI
AC P27051;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN BGI.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1402;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=91224124; PubMed=2026156;
RA Lloberas J., Perez-Pons J.A., Querol E.;
RA "Molecular cloning, expression and nucleotide sequence of the
RA endo-beta-1,3-1,4-D-glucanase gene from Bacillus licheniformis.
RA Predictive structural analyses of the encoded polypeptide.";
RL Eur. J. Biochem. 197;337-343(1991).
RP REVISIONS.
RA Querol E.;

```

```

RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP MUTAGENESIS.
RX MEDLINE=92362869; PubMed=1354172;
RA Planas A., Juncosa M., Lloberas J., Querol E.;
RT "Essential catalytic role of Glu134 in endo-beta-1,3-1,4-D-glucan 4-
RT glucanohydrolase from B. licheniformis as determined by site-directed
RT mutagenesis.";
RL FEBS Lett. 308:141-145(1992).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=94237863; PubMed=8182059;
RA Juncosa M., Pons J., Dot T., Querol E., Planas A.;
RT "Identification of active site carboxylic residues in Bacillus
RT licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase by
RT site-directed mutagenesis.";
RL J. Biol. Chem. 269:14530-14535(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=96063718; PubMed=7589539;
RA Hahn M., Pons J., Planas A., Querol E., Heinemann U.;
RT "Crystal structure of Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-
RT glucanohydrolase at 1.8-A resolution.";
RL FEBS Lett. 374:221-224(1995).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BALEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57279; CAA040547.1; -
DR PIR: S15388; S15388
DR PDB: 1GBG; 07-DEC-95.
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 27
FT CHAIN 28 243 BETA-GLUCANASE.
FT ACT_SITE 134 134 NUCLEOPHILE.
FT ACT_SITE 138 138 PROTON DONOR.
FT DISULFD 61 90
FT MUTAGEN 51 51
FT MUTAGEN 89 89 D->N: 30% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 92 92 D->N: 85% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 95 92 E->Q: 3% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 105 105 E->Q: 50% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 133 133 D->N: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 134 134 E->Q: 0.2% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 136 136 D->N: 0.5% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 138 136 E->Q: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 143 143 D->N: 65% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 160 160 E->Q: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 168 168 D->N: 60% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 179 179 D->N: 80% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 190 190 D->N: 70% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 219 219 D->N: NO CHANGE IN ACTIVITY.
SQ SEQUENCE 243 AA; 27435 MW; 651188D9AAD609A5 CRC64;
Query Match 12.88; Score 233.5; DB 1; Length 243;
Best Local Similarity 33.7%; Pred. No. 7.7e-10;
Matches 58; Conservative 26; Mismatches 77; Indels 11; Gaps 6;
QY 23 TNVSAKDFSGAEIYLTLEEVQYGFKEARMKMAASGVSMFLYONGSEIADGRPWVEVDI 82

```

```

Db      81 TSPYNKFCGSENGSVQYGYGLYEVNMMKPAKNVGISSFFTYGPT---DGTFWDEIDI 137
QY      83 EVLGKNGSGSQSNITITKAGAKTSEKHNAVSPADAFFITGLEMTPNVRVTVGOEV 142
Db      138 EFLKDKDTTKVQFNFTYGVG---NHEKIVNLGFDPAANSYHYAFDPMQNSIKMYVDQ-L 193
QY      143 RKEGCGVSNLTGTQGLRPNLMSSEA-AWVGOFDESKLPLFOFINWVKYK 193
Db      194 KHTATTOIPTGPK--IMNMLNMGAGVDMLGSYN-GVTPLSRSLHWRTK 242

RESULT 8
YGA6_YEAST STANDARD: PRT: 507 AA.
ID YGA6_YEAST STANDARD: PRT: 507 AA.
AC P53301.
DT 01-OCT-1996 (Rel. 34, Last Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 52.8 KDA PROTEIN IN BUB1-HIPI INTERGENIC REGION.
OS YCR189C OR G7553.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=97279231; PubMed=9137339;
RA Attyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nombela C.;
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII.";
RL Yeast 13:357-363(1997).
CC -1- SIMILARITY: SOME, TO YEAST UTR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 272974; CAA97215.1; -
DR EMBL: X99074; CAA67525.1; -
DR HSSP: P23904; IAO.
DR SGD: S0003421; CRH1.
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
KM Hypothetical protein.
FT DOMAIN 63 66 POLY-SER.
FT DOMAIN 301 310 POLY-SER.
FT DOMAIN 345 357 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 467 470 POLY-SER.
FT DOMAIN 507 52757 MM; 7D7B61F57AE942C CRC64;
SQ SEQUENCE 507 AA; 52757 MM; 7D7B61F57AE942C CRC64;

Query Match 12.6%; Score 230; DB 1; Length 507;
Best Local Similarity 26.9%; Pred. No. 3.3e-09;
Matches 87; Conservative 57; Mismatches 134; Indels 46; Gaps 13;

```

```

Db      180 LDWAMDKTWTYLDGESVR-----VLSTNSSEGVPOSPMYLMGICAGDPDAACTIEW 233
QY      172 V-GQFDESKLPLFOFINWVVKYKTPQGGSGDFTLDMWDNDFDTGSRMGKGDWTFDG 230
Db      234 AGGETNNDAPFTWYLEKVTVDYSTGKKTYTGDQGSW-ESTLEADGGTYGRDAQE- 291
QY      231 NRVDTLTGNINYSRGMILALTRKG--QESFNGQVPRDEPAQSSSAPASS--SSVPA 286
Db      292 -----DFVALANGSGTSSSTSTSSSTVSSASSSTVSSSVSTVSSASSSTVSSSTVS 345
QY      287 SSSSVPASSSAFVPPSSSATNA 310
Db      346 SSSSVSSSSSTS---PSSSTATSS 366

RESULT 9
GUB_PAEMA STANDARD: PRT: 237 AA.
ID GUB_PAEMA STANDARD: PRT: 237 AA.
AC P23904;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
OS Paenibacillus macerans (Bacillus macerans).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=44252;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91109712; PubMed=2274030;
RA Borriss R., Buettner K., Maentsaelae P.;
RT "Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
RT homologues to other beta-glucanases.";
RL Mol. Genet. 222:278-283(1990).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=93094208; PubMed=1360982;
RA Hoeft P.B., Condon R., Traeger J.C., McAuliffe J.C., Stone B.A.;
RT "Identification of glutamic acid 105 at the active site of Bacillus
RT amyloliquefaciens 1,3-1,4-beta-D-glucan 4-glucanohydrolase using
RT epoxide-based inhibitors.";
RL J. Biol. Chem. 267:25059-25066(1992).
RN [3]
RP MUTAGENESIS OF GLU-128.
RA Olsen O.;
RT Thesis (1990), University of Aarhus, Denmark.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=93281743; PubMed=8099449;
RA Ketel T., Simon O., Borriss R., Heinemann U.;
RT "Molecular and active-site structure of a Bacillus 1,3-1,4-beta-
RT glucanase.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5287-5291(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=96028129; PubMed=7588726;
RA Hahn M., Ketel T., Heinemann U.;
RT "Crystal and molecular structure at 0.16-nm resolution of the hybrid
RT Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(Al6-M).";
RL Eur. J. Biochem. 232:849-858(1995).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



```

Query Match          10.4%: Score 189.5; DB 1; Length 347;
Best Local Similarity 24.6%: Pred. No. 1.5e-06;
Matches 83; Conservative 49; Mismatches 120; Indels 85; Gaps 15;

QY 28 KDFGAGELTYLLEVOYGFGEARMKMAASGTVMFLYONGSEIADGRPWVEVDIEVLGK 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KNSGTGVLSTRAWYMGKVASRIKSHLAGVVTGILISGAGD-----ELDPEFGA 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 88 NPGSQSNI-----ITGKAGAOQTSKSHNAVSPADQAFHTYGLLEFVPIVYKTVDCQ 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 DLEAQTQFWYESVLYNTNSANISTTD-----TFENYHYELDHDHEDVVTWISIDGV 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 141 EVKRTGGQVGNLNGTOGLRF-----NLW-----SSESA-----AMWG-----QPEESKLP 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 106 VGRILYKNEYTNAT-TQKYQYPTPPSKVDIITWPGCNSTNAPGTITAWSGGETINWASDIS 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 ----LFOFINNVKVKYKYPGOGEGSDYTLDTWDTDFDSDSRMGKGWTFDGNHVDLTD 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 NPGYYAIVNEVNITCYP-----PSDIKKNGTSAVYTSSE-----FLAKIALTD 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 238 KNIV--SFDGMILALTRKKGESFNGQVPRDEPAPQSSSAPASSSSVSPAS----- 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 DEVAMDSDSEG-----SGLDPHKGAT-----TSSTQKSSSSTATSSKTSQSDHSSSTKK 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 ---SSSVASSSSAFVPPSSSSATTAHGMKTPPAAVA 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 SSKTSSTASSSSSSSSSSSSSSSTATKNGDKVSSVS 297

RESULT 13
MERS_ARATH
AC P24806; STANDARD; PRT; 247 AA.
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE MERI-5 PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosida II;
OC Brassicales; Brassicaceae; Arabidopsis.
RX NCBI-TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93005704; Pubmed=1840916;
RA Medford J.I.; Elmer J.S.; Klee H.J.;
RT "Molecular cloning and characterization of genes expressed in shoot
   apical meristems.";
RL Plant Cell 3:359-370(1991).
CC -I- TISSUE SPECIFICITY: EXPRESSED IN SHOOT APICAL MERISTEMS, ALSO
   FOUND IN SEEDLINGS AND MERISTEMS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63166; AAA32028.1; .
DR PIR; J01022; J01022.
DR HSSP; P23904; IAKK.
DR InterPro; IPR000757; .
DR Pfam; PF00722; Glyco_hydro_16; 1
DR SEQUENCE 247 AA; 28295 MW; 831E8441564B45E8 CRC64;

Query Match          8.4%: Score 154; DB 1; Length 247;
Best Local Similarity 26.5%: Pred. NO. 0.00032;

```

```

Matches 53: Conservative 34: Mismatches 71: Indels 42: Gaps 9

QY 6 TAVSALAAVAAAAAALTIVS-----ANDFGSAELYTE-----EVOYGFPE 47
Db 11 TLVAAFVSVAADENTDVNVMNGRGKILLNNGQITLSDLKSSGSGFSQKTEYLFGKID 70
QY 48 ARMKAA--ASGTGVSMPFLYONGSEIADGRPMVEVDIEVLCK---NPGSFOSITITGKAG 102
Db 71 MOKILVPGNSAGTYITFLKSEGS-----TWDEIDFELGMSDDPTTLTNTVYTGCK 124
QY 103 AOKTSEKHAAVSPADAQAFHTYTGLEMTPTNYVMYVDCQ--EYRKTEGGVSNLTGQGLR 160
Db 125 -DKEQGFILWEDPTAN--FHTYSLIMNQRIITLVDTPLRLSLKTMVSLSVSCFQRTKPMR 181
QY 161 F-----NLMSSESAAMV 172
Db 182 MVRFMEROSDMATKRSSWL 201

RESULT 14
GUB_RHOMR STANDARD; PRT; 286 AA.
ID_GUB_RHOMR
AC P45798;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN BGIA.
OS Rhodothermus marinus.
OC Bacteria; CFM group; Rhodothermus group; Rhodothermus.
OX NCBI_TaxID=29549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21 / ITI-378;
RX MEDLINE=95010084; PubMed=7925416;
RA Spillhaert R., Hreggvidsson G.O., Kristjansson J.K.,
RT Eggerston G., Palsdottir A.;
RT Cloning and sequencing of a Rhodothermus marinus gene, bgla, coding
RT for a thermostable beta-glucanase and its expression in Escherichia
RT coli";
RL Eur. J. Biochem. 224:923-930(1994).
CC -1- FUNCTION: SHOWS ACTIVITY ON LICHENAN, BETA-GLUCAN AND LAMINARIN
CC BUT NOT ON CMC CELLULOSE OR XYLAN. THE ENZYME HAS A TEMPERATURE
CC OPTIMUM OF 85 DEGREES CELSIUS AND A PH OPTIMUM OF 7.0.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U04836; AAA60459.1; -.
CC HSSP; P23904; IAKK.
DR InterPro: IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 30
FT CHAIN 1 286
FT ACT_SITE 158 158 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 286 AA; 33145 MW; 7215C3362415191 CRC64;

Query Match 7.3%; Score 133.5; DB 1; Length 286;
Best Local Similarity 23.7%; Pred. No. 0.011;
Matches 44: Conservative 34: Mismatches 85: Indels 23: Gaps 8;
```

```

QY 28 KDFSGAEIYTLLEVO--YGRFEARKMAASGTVSSMFLYONGSEIADGRPW---EVDI 82
DB 103 REYTSARLVTGRKASMTYGRFEIRARLPSCRGTWPATIMLPDRQTSAY-WPONGEIDI 161
QY 83 -EVLGKMP-----GSFOSNITTKAGAKTSEKHAAVSPADQAFHTYGLEWTPYVWTV 137
DB 162 MEHGFNPDPVHGTVHAKVNNHLLGTORGCSIR---VPTARTDPHYVAIETWPEIRMFV 218
QY 138 DCGEVNRKTEGGVSNLGC-----TQGLRFNLMSSSESAAWVG--FDESKLPFOFINW 188
DB 219 DDSLIYRFFPNERLTPEDPDRHWPDPFHLIMNIAVGAMGQGVDPPEAFPAQLVVDY 278
QY 189 VKYKY 194
DB 279 VRYVRR 284

```

```

QY 182 LFOFINWVYKYKT 195
DB 216 FTAYRNFKAIEFS 229

```

Search completed: August 23, 2001, 11:16:32
Job time: 174 sec

```

RESULT 15
BRU1_SOYBN
ID BRU1_SOYBN . STANDARD: PRT: 283 AA.
AC P35654:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE BRASSINOSTEROID-REGULATED PROTEIN BRU1.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-Epicotyl;
RX MEDLINE=94159788; PubMed=8115544;
RA Zurek D.M., Clouse S.D.;
RT "Molecular cloning and characterization of a
RT brassinosteroid-regulated gene from elongating soybean (Glycine max
RT L.) epicotyls."
RL Plant Physiol. 104:161-170(1994).
CC - FUNCTION: POSSIBLE ROLE IN BRASSINOSTEROID-STIMULATED ELONGATION.
CC - SIMILARITY: TO A.THALIANA MERT-5.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L22162; AAA81350.1; -
CC DR HSSP: P23904; IAXK.
CC DR InterPro: IPR000757; -
CC DR Pfam: PF00722; Glyco_hydro_16; 1.
SQ SEQUENCE 283 AA; 32254 MW; C248810EC7835737 CRC64;

```

Query Match 7.0%; Score 128; DB 1; Length 283;
Best Local Similarity 23.7%; Pred. No. 0.025;
Matches 46; Conservative 41; Mismatches 73; Indels 34; Gaps 9;

```

QY 27 AKDFSGAEIYTL-----EVOYGRFEARKMAA--ASGTVSSMFLYONGSEI 71
DB 45 AKTFNGCOLLSLDKVSQSGFSKSKKEYLFGRIIDMLKLVAGNSAGTVAYTL-----S 98
QY 72 ADGRPVVEVDIEVLGK--NPGSFOSNITTKAGAKTSEKHAAVSPADQAFHTYGLEW 128
DB 99 SOGPTDEIDFELGNLSGDPYILHTNINIFTOGKG-NREQOFLMFPDPTNR--FHTYSIIM 155
QY 129 TRPYVWMTVDGQEVKTEGGVSNL--TGTQGLRF--NLMSSSESAAWVG--OFDESKLP 181
DB 156 KPOHILFLVDNPTIRVFNKNAEPLGVPEPRKNQPRISYSLMADMDMATRGGLVKTDMSKAP 215

```


Matches 250: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 2 VSADFSGAEIYLTLEEVYQGFKEAFARMKMAAAGTSSMFLYONGSEIADGRPVEVDIEV 61
 |||||
 Db 25 VSADFGAEIYLTLEEVYQGFKEAFARMKMAAAGTSSMFLYONGSEIADGRPVEVDIEV 84
 |||||
 QY 62 LGKNGSPGOSNIITGKAGAKQTSKHHAVSPADQAEHTYGLEWTPNYVRWTVGQEVK 121
 |||||
 Db 85 LGKNGSPGOSNIITGKAGAKQTSKHHAVSPADQAEHTYGLEWTPNYVRWTVGQEVK 144
 |||||
 QY 122 TEGGVNSLGTGTLRNLNMSSESAAMVQGFDESKLPFQFINNVKYYKTPGCGEGSD 181
 |||||
 Db 145 TEGGVNSLGTGTLRNLNMSSESAAMVQGFDESKLPFQFINNVKYYKTPGCGEGSD 204
 |||||
 QY 182 FTLDWTDNPTFDGSRMGKDMFTDGNRVDLTDKNISRCOMLILALTRKQGESFNGCVP 241
 |||||
 Db 205 FTLDWTDNPTFDGSRMGKDMFTDGNRVDLTDKNISRCOMLILALTRKQGESFNGCVP 264
 |||||
 QY 242 RDEPAPNSSS 252
 |||||
 Db 265 RDEPAPNSSS 275

RESULT 2

GUB_BACBR STANDARD: PRT: 259 AA.

ID GUB_BACBR

AC P37073;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)

DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).

GN Bg1b.

OS Bacillus brevis.

OC Bacteria: Firmicutes: Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Brevibacillus.

OX NCBI_TaxID=1393;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93159752; PubMed=7763386;

RA Louw M.E., Reid S.J., Watson T.G.;

RT "Characterization, cloning and sequencing of a thermostable endo-(1,3-1,4) beta-glucanase-encoding gene from an alkalophilic Bacillus brevis."

RL Appl. Microbiol. Biotechnol. 38:507-513(1993).

-1- FUNCTION: HYDROLYSES B-GLUCANS CONTAINING MIXED BETA-1,3 AND BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; ITS OPTIMAL TEMPERATURE IS NEAR 65-70 DEGREES CELSIUS.

-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.

-1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE SIMILAR TO LICHENASE OF GERMINATING BARLEY.

-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M84339; AAA22265.1; -

DR HSSP: P23904; IAKR.

DR InterPro: IPR000757; -

DR Pfam: PF00722; Glyco_hydro_16; 1.

DR PRINTS: PR00737; GLYDRLASE16.

DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.

KW Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31

FT CHAIN 32 259

FT ACT_SITE 142 146

FT ACT_SITE 146 146

PROTON DONOR (BY SIMILARITY).

SQ SEQUENCE 259 AA; 29960 MW; A63C09F281FF5D13 CRC64;

Query Match 18.5%; Score 266.5; DB 1; Length 259;

Best Local Similarity 36.0%; Pred. No. 36-15;

Matches 63: Conservative 24; Mismatches 75; Indels 13; Gaps 6;

QY 3 SAKDFSGAEIYLTLEEVYQGFKEAFARMKMAAAGTSSMFLYONGSEIADGRPVEVDIEV 62
 |||||
 Db 90 SARYKAGELRTNDFYHYGLFEVSMKPAKVEGYSSFFTY-TGEMDMDGDPMEIDIEFL 148
 |||||
 QY 63 GRNPGSPGOSNIITGKAGAKQTSKHHAVSPADQAEHTYGLEWTPNYVRWTVGQEVKRT 122
 |||||
 Db 149 GKDTTRIQFNFYFTNGVGG---NEFYDILGFDASESFPNTYAFEREDSITWYVNGEAVHTA 205
 |||||
 QY 123 EGGGVNSLGTGTLRNLNMSSESA-AMVQGFDESKLPFQFINNVKYYKTPGCG 175
 |||||
 Db 206 ---TENIPQTPKIMNLMFPGVGDGTVFGDNTPVSYDWW---RYTPIQ 253
 |||||

RESULT 3

GUB_BACSU STANDARD: PRT: 242 AA.

ID GUB_BACSU

AC F04957;

DT 13-AUG-1987 (Rel. 05, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)

DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).

GN BGLS OR BGL OR LIC5 OR N15B.

OS Bacillus subtilis.

OC Bacteria: Firmicutes: Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=C120;

RX MEDLINE=84272222; PubMed=6087283;

RA Murphy N., McConnell D.J., Cantwell B.A.;

RT "The DNA sequence of the gene and genetic control sites for the excreted B. subtilis enzyme beta-glucanase."

RL Nucleic Acids Res. 12:5355-5367(1984).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=HL-25;

RA Tezuka H., Yuuki T., Yabuchi S.;

RT "Construction of a beta-glucanase hyperproducing Bacillus subtilis using the cloned beta-glucanase gene and a multi-copy plasmid."

RL Agric. Biol. Chem. 53:2335-2339(1989).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / BGSJAL;

RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;

RT Submitted (FEB-1995) to the EMBL/GenBank/DBD databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=95219081; PubMed=7704256;

RA Wolf M., Geetz A., Simon O., Boriss R.;

RT "Genes encoding xylan and beta-glucan hydrolizing enzymes in Bacillus subtilis: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase."

RL Microbiology 141:281-290(1995).

RN [5]

RP SEQUENCE OF 1-6 FROM N.A.

RC STRAIN=BR151;

RX MEDLINE=96178961; PubMed=8606172;

RA Schneitz K., Stuelke J., Gertz S., Krueger S., Krieg M., Hecker M., Rak B.;

RT "Bact, a Bacillus subtilis transcriptional antiterminator protein of the BglG family."

RL J. Bacteriol. 178:1971-1979(1996).

RN [6]

```

RA  SEQUENCE OF 29-63.
RA  YUKU T., Tezuka H., Yabuuchi S.;
RT  "Purification and some properties of two enzymes from a beta-glucanase
RL  hyperproducing strain, Bacillus subtilis HL-25."
RT  Agric. Biol. Chem. 53:2341-2346(1989).
CC  -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC  IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC  -1- SUBCELLULAR LOCATION: SECRETED.
CC  -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC  SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC  -1- MISCELLANEOUS: THE SEQUENCE OF STRAIN 168 IS SHOWN.
CC  -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: X00754; CAA25328.1; -
DR  EMBL: D00518; BAA00405.1; -
DR  EMBL: D83026; BAA11697.1; -
DR  EMBL: Z46862; CAA86922.1; -
DR  EMBL: Z28340; CAA82195.1; -
DR  EMBL: Z29124; CAA15943.1; -
DR  PIR: A22914; LXBS
DR  PIR: JU0110; JU0110.
DR  HSSP: P27051; IGBG.
DR  Subtilisin; Bg10476; bg1s.
DR  InterPro: IPR000757; -.
DR  Pfam: PF00722; Glyco_hydro_16; 1.
DR  PRINTS: PR00737; GLHYDRASE16.
DR  PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
DR  KMW: HydroLase; Glycosidase; Signal.
FT  CHAIN 1 28
FT  MOD_RES 29 242 BETA-GLUCANASE.
FT  ACT_SITE 133 133 NUCLEOPHILE (BY SIMILARITY).
FT  ACT_SITE 137 137 PROTON DONOR (BY SIMILARITY).
FT  DISULFID 60 89 BY SIMILARITY.
FT  VARIANT 24 24 A -> S (IN STRAIN HL-25).
FT  VARIANT 83 83 A -> S (IN STRAIN HL-25).
FT  VARIANT 204 204 P -> L (IN STRAIN C120).
SQ  SEQUENCE 242 AA; 27268 MW; 45958DEA70F22B29 CRC64;
Query Match 17.1%; Score 245.5; DB 1; Length 242;
Best Local Similarity 35.8%; Pred. No. 1.5e-13;
Matches 59; Conservative 24; Mismatches 71; Indels 11; Gaps 6;
OY 7 FSGALYLYLEVOYGFKFAKMAAAGTSSMFLYONGSELAQGRPVVEVDIEVLGNP 66
DB 87 FDCGNSRSVQYGLYGLYEVKRAKMTGIVSSFFTYGTPT--DQTPWDEIDIEFLGKDT 143
OY 67 GSFOSNITGAGAGAKTSEKHNHVAAPADQAFHTYGLTEKTPYVYVATQGEVREKTEBGO 126
DB 144 TKVQPNYNTTNGAG--NNEKIVDLGFDAANAHYTAFFWOPNSIKWYVDGO LKHTATNQ 199
OY 127 VSNLTGTOGLRFENLMSSEA-AWGGPFDESKLPLEFQITNWKVVK 170
DB 200 IPTTPGK--IMMNLNMGTVGDEWLGSYNGVN-PLVAHYDWMRYTK 241
RESULT 4
XYND_RUMFL STANDARD; PRT; 802 AA.
AC 053317;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE XYLANASE/BETA-GLUCANASE PRECURSOR [INCLUDES: ENDO-1,4-BETA-XYLANASE

```

```

DE (EC 3.2.1.8) (XYLANASE); (ENDO-BETA-1,3-1,4 GLUCANASE) (EC 3.2.1.73)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE)].
DE XYND.
GN Ruminococcus flavefaciens.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Ruminococcus.
OX NCBI_Taxid=1265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17;
RX MEDLINE=93259938; PubMed=8491715;
RA Flint H.J., Martin J., McPherson C.A., Daniel A.S., Zhang J.-X.;
RT "A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-
RT flavefaciens."
RT flavefaciens."
RL J. Bacteriol. 175:2943-2951(1993).
CC -1- FUNCTION: CONTAINS TWO CATALYTIC DOMAINS WITH XYLANASE AND ENDO-
CC BETA-1,3-1,4 GLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC LINKAGES IN XYLAN.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- PATHWAY: XYLAN DEGRADATION.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC G (FAMILY 11 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 16 OF
CC GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: S61204; AAB26620.1; -
DR  HSSP: P23904; IAKK.
DR  InterPro: IPR000757; -.
DR  InterPro: IPR001137; -.
DR  Pfam: PF02018; CBD_6; 1.
DR  Pfam: PF00457; Glyco_hydro_11; 1.
DR  Pfam: PF00722; Glyco_hydro_16; 1.
DR  PRINTS: PR00737; GLHYDRASE16.
DR  PRINTS: PR00911; GLHYDRASE11.
DR  PROSITE: PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR  PROSITE: PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR  PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
DR  KMW: Xylan degradation; HydroLase; Glycosidase; Signal;
KW Multifunctional enzyme.
FT  SIGNAL 1 31
FT  CHAIN 32 802
FT  DOMAIN 32 244 XYLANASE/BETA-GLUCANASE.
FT  DOMAIN 245 523 A (XYLANASE).
FT  DOMAIN 524 555 B.
FT  DOMAIN 556 802 LINKER.
FT  ACT_SITE 124 124 C (BETA-GLUCANASE).
FT  ACT_SITE 124 124 NUCLEOPHILE (BY SIMILARITY).
FT  ACT_SITE 226 226 PROTON DONOR (BY SIMILARITY).
FT  DOMAIN 684 684 NUCLEOPHILE (BY SIMILARITY).
FT  DOMAIN 524 529 POLY-THR.
FT  DOMAIN 532 543 POLY-THR.
FT  DOMAIN 546 553 POLY-THR.
SQ  SEQUENCE 802 AA; 89091 MW; 2880A689647284AF CRC64;
Query Match 16.9%; Score 242.5; DB 1; Length 802;
Best Local Similarity 34.9%; Pred. No. 1.1e-12;
Matches 58; Conservative 21; Mismatches 74; Indels 13; Gaps 6;
OY 7 FSGALYLYLEVOYGFKFAKMAAAGTSSMFLYONGSELAQGRPVVEVDIEVLGNP 66
DB 634 YSGGEFRNNFHYGYECSSQAMKNDGVSSFFTYGPS--DDNPWDEIDIEFLGKNT 690

```

OY 67 GSFQSNITTKAGAKTSEKHHA VSPAQAQFHTYGLEMTPNVYRWTVDOEVRKTEGQ 126
 DB 691 TOYQFNNTGOGKH---EKYLDGFDSSAEAHYTGFDWQPNYIAWVDREYRA---- 743
 OY 127 VSNLTGOG-LRFENLWSESA-AWVGQFDESKLPLFOFINWVYK 170
 DB 744 TODIPKPKIMNAMPGLTVDDMLKAFN-GRPTPLAHYQWVYK 788

RESULT 5

GUB_BACAM STANDARD: PRT; 239 AA.
 AC P07980;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLA.
 OS Bacillus amyloliquefaciens.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1390;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BE 20/78;
 RX MEDLINE=87192007; PubMed=3106158;
 RA Hofemeister J., Kurtz A., Borris R., Knowles J.;
 RT "the beta-glucanase gene from Bacillus amyloliquefaciens shows
 extensive homology with that of Bacillus subtilis.";
 RL Gene 49:177-187(1986).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M15674; AAA87323.1; -
 CC DR PIR: A29091; A29091.
 CC DR HSSP: P27051; 1GBG.
 CC DR InterPro: IPR000757; -
 CC DR Pfam: PF00722; GLYCO_hydro.16; 1.
 CC DR PRINTS: PR00737; GLHYDRLASE16.
 CC DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 CC KM Hydrolyase; Glycosidase; Signal.
 CC FT SIGNAL 1 25
 CC FT CHAIN 26 239 BETA-GLUCANASE.
 CC FT ACT_SITE 134 134 NUCLEOPHILE (BY SIMILARITY).
 CC FT DISULFID 57 86 BY SIMILARITY.
 CC SQ SEQUENCE 239 AA; 26928 MW; A76A64268A7AA0B CRC64;

Query Match 16.6%; Score 238.5; DB 1; Length 239;
 Best Local Similarity 34.3%; Pred. No. 5.7e-13;
 Matches 58; Conservative 26; Mismatches 74; Indels 11; Gaps 6;

OY 3 SAKDFSGAELTYLEVOYKFEARMMAAGSVSSMFLYONGSETADGRWVEVDIEVL 62
 DB 80 SYNKFCCGGERSVQTYGYGLYEVKMPAKNTGIVSSFFTYTGPT---GTPMDEIDIEFL 136
 OY 63 GRNPGSFQSNITTKAGAKTSEKHHA VSPAQAQFHTYGLEMTPNVYRWTVDOEVRK 122
 DB 137 GKDTTVQFENYNTGANG---NHKEFADLDGDAANAHYTAFAEDQPNYSIKWYVDOG-LKRT 192
 OY 123 EGGQVSNLTGOG-LRFENLWSESA-AWVGQFDESKLPLFOFINWVYK 170

DB 193 ATTQIDPAAPK--IMNIMMGTCVDDMLGSGYCNV-PIYAHYDMRYRK 238

RESULT 6

GUB_CLOTM STANDARD: PRT; 334 AA.
 AC P29716; P37074;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LAMINARINASE).
 GN LICB OR LAM1.
 OS Clostridium thermocellum.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 27405 / DSM 1237;
 RX MEDLINE=92155194; PubMed=1740123;
 RA Schimming S., Schwarz W.H., Staudenbauer W.L.;
 RT "Structure of the Clostridium thermocellum gene licB and the encoded
 RT beta-1,3-1,4-glucanase. A catalytic region homologous to Bacillus
 RT lichenases joined to the reiterated domain of clostridial
 RT cellulases.";
 RL Eur. J. Biochem. 204:13-19(1992).
 RN [2]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RC STRAIN-E7;
 RX MEDLINE=92095946; PubMed=1755832;
 RA Zverlov V.V., Laptev D.A., Tishkov V.I., Velikodvorskaja G.A.;
 RT "Nucleotide sequence of the Clostridium thermocellum laminarinase
 RT gene.";
 RL Biochem. Biophys. Res. Commun. 181:507-512(1991).
 RN [3]
 RP REVISIONS.
 RA Zverlov V.V.;
 RC Submitted (Aug-1993) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RA Zverlov V.V., Velikodvorskaja G.A.;
 RC STRAIN-E7;
 RT "Cloning the Clostridium thermocellum thermostable laminarinase gene
 RT in Escherichia coli; the properties of the enzyme thus produced.";
 RL Biotechnol. Lett. 12:811-816(1990).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- SUBUNIT: MAY FORM PART OF A MULTIENTZYME COMPLEX (CELLULOSE).
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMS. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X63355; CAA44959.1; -
 CC EMBL: X58392; CAA41281.1; -
 CC DR PIR: S23498; S23498.
 CC DR PIR: JS0611; JS0611.
 CC DR PIR: S18726; S18726.
 CC DR HSSP: P23904; 1AJK.
 CC DR InterPro: IPR000757; -
 CC DR InterPro: IPR002048; -
 CC DR InterPro: IPR002105; -
 CC DR Pfam: PF00404; Dockerin_1; 2.

```

DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
DR PROSITE; PS00018; EF_HAND_UNKNOWN_2.
DR PROSITE; PS00448; CLUS_CELULOSOME_RPT; 2.
KW Hydrolyase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 334
FT ACT_SITE 136 136 BETA-GLUCANASE.
FT ACT_SITE 140 140 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 252 269 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 273 331 PRO/THR-RICH (LINER).
FT REPEAT 273 296 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 308 331 1.
FT REPEAT 308 331 2.
FT CONFLICT 304 334 OSVADVNRDRGRIDSTDTMLKRYLIRAPSL ->
PODCGHRDVRVDSGSK (IN REF. 2 AND 3).
SO SEQUENCE 334 AA; 37897 MW; 0837564E9726F281 CRC64;

Query Match 16.5%; Score 237; DB 1; Length 334;
Best Local Similarity 30.9%; Pred. No. 1,le-12;
Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;

OY 5 KDFSGAELTYLEVO-----YGFKFAKMAAASGVSSMFLYONGSEIADGRPEVND 58
DB 82 REYGGSYPKSGEYRTKSGFSGYGEYVPMKAANKYIVSFEFTYGPS---DNNPWEID 138
OY 59 IEVIGKNGNSQSNITITKAGAKTSEKHNAVSPADAFTHTGLEMTPNVYRWTVDDOE 118
DB 139 IEFGKDTTKVQFMWYKNGVSG--NEXLHMGFDASODFHTYGEWMPDYIDFYVDGK 195
OY 119 VRTKEGGGVSNLTGTQG-LRFNLMSESA-AWVGQFDESKLPFLFOFMWVKY----- 169
DB 196 VYR-----GTRNIPVTPKAIMNMLWPGIGVDMLGRYD-GRPLQAEYEVKTYRPGVPOD 250
OY 170 -----KTRP-----GQGE-GGSDFT-----LDWTNDFPFDSGRWKG 201
DB 251 NPTPTPTAPSTPTPNPLKGDVNGDGVNSDYSLEKRYLLRVIDREFPVDDQS--VA 307
OY 202 DMTFDCGNRVDLTDKNIYSR 220
DB 308 DVNRDG-RIDSTDTMLKR 325

RESULT 7
GUB_BACLI STANDARD: PRT: 243 AA.
ID GUB_BACLI
AC P27051;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN Bgl.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OC NCBI_TaxID=1402;
RX NCBI_TaxID=1402;
RX SEQUENCE FROM N.A.
RX MEDLINE=91224124; PubMed=2026156;
RA Lloberas J., Perez-Pons J.A., Querol E.;
RA "Molecular cloning, expression and nucleotide sequence of the
RA endo-beta-1,3-1,4-D-glucanase gene from Bacillus licheniformis.
RA Predictive structural analyses of the encoded polypeptide.";
RL Eur. J. Biochem. 197;337-343(1991).
RN [12]
RN REVISIONS.
RA Querol E.;
RN Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RX MUTAGENESIS.
RX MEDLINE=92362869; PubMed=1354172;

```

```

RA Planas A., Juncosa M., Lloberas J., Querol E.;
RT "Essential catalytic role of Glu134 in endo-beta-1,3-1,4-D-glucan 4-
RT glucanohydrolase from B. licheniformis as determined by site-directed
RT mutagenesis.";
RL FEBS Lett. 308:141-145(1992).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=94237863; PubMed=8182059;
RA Juncosa M., Pons J., Dot T., Querol E., Planas A.;
RT "Identification of active site carboxylic residues in Bacillus
RT licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase by
RT site-directed mutagenesis.";
RL J. Biol. Chem. 269:14530-14535(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=96063718; PubMed=7589539;
RA Hahn M., Pons J., Planas A., Querol E., Heinemann U.;
RT "Crystal structure of Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-
RT glucanohydrolase at 1.8-A resolution.";
RL FEBS Lett. 374:221-224(1995).
CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -I- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -I- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X57279; CAA40547.1; -.
DR PIR; S15368; S15368.
DR PDB; 1GBG; 07-DEC-95.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
KW PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolyase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 27
FT CHAIN 28 243
FT ACT_SITE 134 134 BETA-GLUCANASE.
FT ACT_SITE 138 138 NUCLEOPHILE.
FT DISULFID 61 90 PROTON DONOR.
FT MUTAGEN 51 51
FT MUTAGEN 89 89
FT MUTAGEN 92 92
FT MUTAGEN 105 105
FT MUTAGEN 133 133
FT MUTAGEN 134 134
FT MUTAGEN 136 136
FT MUTAGEN 138 138
FT MUTAGEN 143 143
FT MUTAGEN 160 160
FT MUTAGEN 168 168
FT MUTAGEN 179 179
FT MUTAGEN 190 190
FT MUTAGEN 219 219
SO SEQUENCE 243 AA; 27435 MW; 651188D9AD609A5 CRC64;

Query Match 15.9%; Score 229.5; DB 1; Length 243;
Best Local Similarity 33.7%; Pred. No. 3,3e-12;
Matches 57; Conservative 25; Mismatches 76; Indels 11; Gaps 6;

OY 3 SAKDFSGAELTYLEVOYGFKFAKMAAASGVSSMFLYONGSEIADGRPEVNDIEVL 62
DB 84 SYNKFDCGNSVOTYTGLETVNMPKAKNGIVSFEFTYGPV---DGTWDEIDIEFL 140
OY 63 GKNPDSFOSNITITKAGAKTSEKHNAVSPADAFTHTGLEMTPNVYRWTVDDOEVRKT 122

```

```

DB 141 GKDTKVOFNYYTNGVG---NHEKIVNLGFDANSHYHTAFDQWPNISIKWYVDCO-LKHT 196
OY 123 EGGGVSNLTGNGGLRFLNLMSESA-AWVGOPDESKLPFOFINWVYKXK 170
DB 197 ATTQIIPOTPGK--IMNMLMGAGVDENLGSYN-GVTPLSNLSHWRYTK 242

RESULT 8
GUB_PAEMO STANDARD; PRT; 237 AA.
ID GUB_PAEMO
AC P23904;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
OS Paenibacillus macerans (Bacillus macerans).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=44252;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91109712; PubMed=2274030;
RA Borriss R., Buettner K., Maentzsaellae P.;
RT "Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
RT homologues to other beta-glucanases."
RL Mol. Gen. Genet. 222:278-283(1990).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=93094208; PubMed=1360982;
RA Hoel P.B., Condon R., Traeger J.C., McAuiffie J.C., Stone B.A.;
RT "Identification of glutamic acid 105 at the active site of Bacillus
RT amyloliquefaciens 1,3-1,4-beta-D-glucan 4-glucanohydrolase using
RT epoxide-based inhibitors."
RL J. Biol. Chem. 267:25059-25066(1992).
RN [3]
RP MUTAGENESIS OF GLU-128.
RA Olsen O.;
RL Thesis (1990), University of Aarhus, Denmark.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=93281743; PubMed=8099449;
RA Keitel T., Simon O., Borriss R., Heinemann U.;
RT "Molecular and active-site structure of a Bacillus 1,3-1,4-beta-
RT glucanase."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5287-5291(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=96028129; PubMed=7588726;
RA Hahn M., Keitel T., Heinemann U.;
RT "Crystal and molecular structure at 0.16-nm resolution of the hybrid
RT Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(al6-w)."
RL Eur. J. Biochem. 232:849-858(1995).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
CC EMBL: X55959; CAA39426.1; -
CC PIR: S11927; S11927
CC PDB: 2AYH; 31-MAR-95.
CC PDB: 1BYH; 31-OCT-93.
CC PDB: 1CPM; 22-JUN-94.

```

```

DR PDB; 1CPM; 22-JUN-94.
DR PDB; 1MAC; 27-FEB-95.
DR PDB; 1AJK; 06-MAY-98.
DR PDB; 1AJO; 06-MAY-98.
DR InterPro: IPR000757;
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal; 3d-structure.
FT SIGNAL 1 23
FT CHAIN 24 237
FT ACT_SITE 128 128
FT ACT_SITE 132 132
FT DISULFID 55 84
FT MOTAGEN 128 128
FT STRAND 26 26
FT STRAND 29 31
FT STRAND 41 43
FT TURN 51 52
FT STRAND 53 53
FT STRAND 56 57
FT STRAND 59 61
FT HELIX 62 64
FT STRAND 66 67
FT TURN 70 78
FT STRAND 79 80
FT TURN 81 89
FT STRAND 93 93
FT STRAND 96 103
FT TURN 108 109
FT STRAND 110 118
FT HELIX 120 122
FT TURN 123 123
FT STRAND 127 134
FT TURN 135 136
FT STRAND 137 147
FT TURN 148 149
FT STRAND 150 150
FT STRAND 155 158
FT TURN 163 164
FT STRAND 168 174
FT STRAND 179 183
FT TURN 184 185
FT STRAND 186 191
FT STRAND 200 209
FT TURN 213 216
FT STRAND 225 236
SQ SEQUENCE 237 AA; 26589 MW; 436EABCDFFC87781 CRC64;

Query Match 15.7%; Score 226; DB 1; Length 237;
Best local Similarity 35.1%; Pred. No. 6.2e-12;
Matches 59; Conservative 23; Mismatches 70; Indels 16; Gaps 8;

OY 7 FSGAEIYTLLEVOYGRFEARKMAAAGTVSSMFLYQNGSEIADGRPWVEVDIEVLGNP 66
DB 82 FDCAEKSTNIYGYGLYEVKMPAKMNTGIVSFEYTG---AHGTQWDEIDIEFLGKDT 138
OY 67 GSPQSNITTKACAOQTSKHNHVAVPAADOAEFTTYGLETPNTVKNRTVDCQEVKRTBGG 126
DB 139 TKQOFNYYTNGVG---ELVISLGFDAKSGFTYAFDQWPGYIKWYVDB-VLKHT---A 191
OY 127 VSNLTGNGGLRFLNLMSESA-AWVGOPDESKLPFOFINWVYKXK 172
DB 192 TANIPSTPGKIMNMLMGAGVDENLGSYN-PLTAEYDMV---KYT 235

RESULT 9
GUB_PAEMO STANDARD; PRT; 238 AA.
ID GUB_PAEMO
AC P45797;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```

DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 GN GLOB.
 OS Paenibacillus polymyxa (Bacillus polymyxa).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Paenibacillus.
 RX NCBI_TaxID=1406;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 842;
 RX MEDLINE=92041687; PubMed=1938968;
 RA Gosalbes M.D., Perez-Gonzalez J.A., Gonzalez R., Navarro A.;
 RT "Two beta-glucanase genes are clustered in Bacillus polymyxa:
 RT molecular cloning, expression, and sequence analysis of genes
 RT encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
 RL J. Bacteriol. 173:7705-7710(1991).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X57094; CAA40379.1; -
 DR HSSP; P23904; ICPM.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS; PR00737; GLHIDRLASE16.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 238
 FT ACT_SITE 129 129 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 133 133 PROTON DONOR (BY SIMILARITY).
 FT DISULFID 56 85 BY SIMILARITY.
 SQ SEQUENCE 238 AA; 26919 MW; C0CFB4E5D40E8C CRC64;
 Query Match 15.6%; Score 225; DB 1; Length 238;
 Best Local Similarity 34.5%; Pred. No. 7.5e-12;
 Matches 58; Conservative 23; Mismatches 71; Indels 16; Gaps 8;
 Oy 7 FSGAEVLYLEVOYGKFEARKMAAASGVSSMFLYONGSEIADGRPVVEVIEVLGNP 66
 Db 83 FDCGEYRTNNYGYGLYEVSKPAKNTIVSSFFTYTSPH---GTQWDEIDIEFLGKDT 139
 Oy 67 GSPFSNITTKRAGAKTSEKHNHAAVSPADQAHTHTGLEMTPRYVWYWDGQEVKRTGEGQ 126
 Db 140 TKYQFNNTYTVGVGH---EKIINIGFDASTSEPHYAFPMQGYIKWYVDG-VLKHT---A 192
 Oy 127 VSNLTGTGOG-LRFNLMSSEA-AWVGQFDESKLLPQFINNVKYYKKT 172
 Db 193 TTNIPTSPFGKIMMNLNMGTVDSWISGYSYGAN-PLYAEYDMV---KYT 236
 RESULT 10
 EXOK_RHIME STANDARD; PRT; 269 AA.
 AC P33693;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ENDO-1,3-1,4-BETA-GLUCANASE EXOK PRECURSOR (EC 3.2.1.-) (SUCCINOGLYCAN
 DE BIOSYNTHESIS PROTEIN EXOK).
 GN EXOK.

OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSym (megaplasmid 2).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 RX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=94042869; PubMed=8226645;
 RA Glucksmann M.A., Reuber T.L., Walker G.C.;
 RT "Family of glycosyl transferases needed for the synthesis of
 RT succinoglycan by Rhizobium meliloti.";
 RL J. Bacteriol. 175:7033-7044(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RCR2011 / SU47;
 RX MEDLINE=93241147; PubMed=8479421.
 RA Becker A., Kleckmann A., Arnold W., Puhler A.;
 RT "Analysis of the Rhizobium meliloti exoH/exoK/exoL fragment: ExoK
 RT shows homology to excreted endo-beta-1,3-1,4-glucanases and ExoH
 RT resembles membrane proteins.";
 RL Mol. Gen. Genet. 238:145-154(1993).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98226741; PubMed=9560202;
 RA York G.M., Walker G.C.;
 RT "The Rhizobium meliloti ExoK and ExoH glucanases specifically
 RT depolymerize nascent succinoglycan chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4912-4917(1998).
 CC -1- FUNCTION: CLEAVES HIGH MOLECULAR WEIGHT SUCCINOGLYCAN TO YIELD LAM
 CC SUCCINOGLYCAN. DYNAMICALLY REGULATE THE MOLECULAR WEIGHT
 CC DISTRIBUTION OF SUCCINOGLYCAN BY CLEAVING NASCENT SUCCINOGLYCAN
 CC ONLY DURING A LIMITED PERIOD AFTER ITS SYNTHESIS, PERHAPS BEFORE
 CC IT UNDERGOES A TIME-DEPENDENT CHANGE IN ITS CONFORMATION OR
 CC AGGREGATION STATE.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L20758; AAA16048.1; -
 DR EMBL; Z17219; CAA78927.1; -
 DR HSSP; P23904; IAKJ.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Exopolysaccharide synthesis; Glycosidase; Hydrolase; Signal; Plasmid.
 FT SIGNAL 1 27
 FT CHAIN 28 269
 FT CONFLICT 93 102 ACCEIOTRKR -> LRNPADQG (IN REF. 2).
 SQ SEQUENCE 269 AA; 30083 MW; 073CC7ED65EF2611 CRC64;
 Query Match 13.7%; Score 197; DB 1; Length 269;
 Best Local Similarity 32.5%; Pred. No. 1.8e-09;
 Matches 50; Conservative 31; Mismatches 61; Indels 12; Gaps 6;
 Oy 2 VSADFSGAEVLYLEVOYGKFEARKMAAASGVSSMFLYONGSEIADGRPVVEVIEVL 61
 Db 87 VKERNFACGEIQTFRKRFYGYTEARIKAADSGINSAFETYICP---ADKKRPHDEIDREV 143
 Oy 62 LKKNRGSQSN-IITGRKAGAKTSEKHNHAAVSPADQAHTHTGLEMTPRYVWYWDGQEV 120
 Db 144 LKNTAKVQIQYQVSAKSGNFEIAD---VPGANGQNDYAFYWKRIIRYVNGELV 199
 Oy 121 K-TEGGQVSNLTGTGGLRFLNLSSEAA-WVGQF 152


```

RL Submitted (FEB-1993) to the EMBL/Genbank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Borstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-B-6441;
RX MEDLINE-94016558; PubMed-8411151;
RA Melnick L., Sherman F.;
RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
RT of Saccharomyces cerevisiae share a common ancestry."
RL J. Mol. Biol. 233:372-388(1993).
CC -1- SIMILARITY: SOME, TO YEAST YGR189C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL, U18779; AAB65002.1; ALT_INT.
DR EMBL, L22173; AAA34941.1;
DR EMBL, S65964; AAD13975.1;
DR EMBL, S66130; AAB28444.1;
DR PIR, S30839; S30839.
DR HSSP, P23904; 1AJO.
DR SGO, S0000766; UTR2.
DR InterPro: IPR000757;
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR DOMAIN 234 322
FT SER-RICH.
FT DOMAIN 269 283
FT POLY-SER.
FT CONFLICT 10 10 L->V (IN REF. 3).
FT CONFLICT 171 171 A->R (IN REF. 3).
FT CONFLICT 234 234 S->C (IN REF. 3).
SQ SEQUENCE 347 AA; 36692 MW; 1E8AFB862C4BB328 CRC64;

Query Match 10.2%; Score 147; DB 1; Length 347;
Best Local Similarity 25.1%; Pred. No. 3.6e-05;
Matches 50; Conservative 30; Mismatches 71; Indels 48; Gaps 9;

OY 5 KDFSGAEIYTLIEVQ--YKFEARMKMAASGVSSMFLYONGSEIADGRWVEVDIEVLGK 64
DB 3 KNSGTVLSTRVAVWGKVASRIKTSHLGVLGYFLSGAD-----ELDYEFVGA 54
OY 65 NFGSFOSNI-----ITGKAGAOKTSEKHAIVSPAADAFTYGLGEMTPNVYRWTVDO 117
DB 55 DLEFQINPYWESVLYNTSANSISTD-----TFENYITFELDMEDYVTSIDYV 105
OY 118 EVRKTEGGQVSNLTGQGLRF-----NLN--SSESA--AWG--QFDESKLP 158
DB 106 VGRILYKNETYNAT--TKQYQYQPTPSKVIDSIMPQGNSTNAGDTIAMSCEINWDASDIS 164
OY 159 ----LQFQINWKKVKKYTP 173
DB 165 NPGYYIAIVNEVNYTCYDP 183

RESULT 14
GUB_RHOMR STANDARD; PRT; 286 AA.
AC P45798;
DT 01-NOV-1995 (Rel. 32, Created)

```

```

DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN BGLA.
OS Rhodothermus marinus.
OC Bacteria; CFb group; Rhodothermus group; Rhodothermus.
OX NCBI_TaxId=29549;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-21 / ITI-378;
RX MEDLINE-95010084; PubMed-7925416;
RA Spilliaert R., Hreggvidsson G.O., Kristjansson J.K.,
RA Eggerissson G., Paldottir A.;
RT "Cloning and sequencing of a Rhodothermus marinus gene, bglA, coding
RT for a thermostable beta-glucanase and its expression in Escherichia
RT coli."
RL Eur. J. Biochem. 224:923-930(1994).
CC -1- FUNCTION: SHOWS ACTIVITY ON LICHENAN, BETA-GLUCAN AND LAMINARIN
CC BUT NOT ON CMC, CELLULOSE OR XILAN. THE ENZYME HAS A TEMPERATURE
CC OPTIMUM OF 85 DEGREES CELSIUS AND A PH OPTIMUM OF 7.0.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
DR EMBL, U04836; AAA60459.1;
DR HSSP, P23904; 1AJK.
DR InterPro: IPR000757;
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolyase; Glycosidase; Signal.
FT SIGNAL 1 30
FT CHAIN 31 286
FT ACT_SITE 158 158 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 286 AA; 33145 MW; 7215C33624135191 CRC64;

Query Match 9.3%; Score 133.5; DB 1; Length 286;
Best Local Similarity 23.7%; Pred. No. 0.00037;
Matches 44; Conservative 34; Mismatches 85; Indels 23; Gaps 8;

OY 5 KDFSGAEIYTLIEVQ--YKFEARMKMAASGVSSMFLYONGSEIADGRWVEVDIEVLGK 59
DB 103 REYTSARLVTRGKASWTVYGRFELRLPSGRGTWPAIMLPDQYGYSAV--WDNGEIDI 161
OY 60 -EVLGNP-----GSFOSNIITGKAGAOKTSEKHAIVSPAADAFTYGLGEMTPNVYRWTV 114
DB 162 MEHVGFNPDPVAVGTVTKAYNLHLGTQROGGSIR--VPAARDHFHYALAEWPEELRWTV 218
OY 115 DGOEVKTEGGQVSNLTG-----TQGLRFNLWSSESAAMWGQ--FDESKLPQFQINW 165
DB 219 DDLVYRFPNERLTDEADRWHPFDQPHLLIMNIAVGGAMGQGVDPDEAFPAQLVVDY 278
OY 166 VKVYKY 171
DB 279 VRVYRW 284

RESULT 15
BRUL_SOYBN STANDARD; PRT; 283 AA.
AC P35694;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)

```

DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE BRASSINOSTEROID-REGULATED PROTEIN BRU1.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;
 OC Fabales; Fabaceae; Papilionoideae; Glycine.
 ON NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Epicotyl:
 RX MEDLINE=94159788; PubMed=8115544;
 RA Zurek D.M., Clouse S.D.;
 RT "Molecular cloning and characterization of a
 RT brassinosteroid-regulated gene from elongating soybean (Glycine max
 RT L.) epicotyls."
 RL Plant Physiol. 104:161-170(1994).
 CC -!- FUNCTION: POSSIBLE ROLE IN BRASSINOSTEROID-STIMULATED ELONGATION.
 CC -!- SIMILARITY: TO A.THALIANA MERI-5.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L23162; AAA81350.1; -
 DR HSSP: P23904; JAK.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR SEQUENCE 283 AA; 32254 MW; C248810EC7835737 CRC64;

Query Match 8.98; Score 128; DB 1; Length 283;

Best Local Similarity 23.78; Pred. No. 0.0011; Mismatches 73; Indels 34; Gaps 9;

Matches 46; Conservative 41; Indels 34; Gaps 9;
 QY 4 AKDEGAELVTL-----EEVOYGFAPRKMAA--ASGVSSMFLYONGSEI 48
 DB 45 AKINFGQLSLSDKYSKSGFKSKKEYLFRIDMQLKLVAGNSAGIVTAYTL-----S 98
 QY 49 ADGRPWVEVDIEVLGK--NPGSFOSNITGKAGAKQKTSKHNHNVSPAADQAFHTYGLEW 105
 DB 99 .SQGPTHEIDIEFEFLGNSGDPYILHTNIFTOGKG-NREQPFYLMEDPTRN--FHTYSIIW 155
 QY 106 TPNVYRRTVDGQEVKRTGEGVSNL--TGQGLRF--NLMSSESAANVG---QFDESKLP 158
 DB 156 KPOHITFLVDTPTRVFNKNAEPLGVPPEPKNOPMRYSSLMNADMDATRGGLVKTDMWSKAP 215
 QY 159 LFOFINNVKVKYKT 172
 DB 216 FTAVYRNFKALIEFS 229

Search completed: August 23, 2001, 11:16:31
 Job time: 173 sec

A:Molecule type:: DNA
A:Residues: 1-242 <KUN>
A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CAB15943.1; PID:g2636453
A:Experimental source: strain 168
R:Murphy, N.; McConnell, D.J.; Cantwell, B.A.
Nucleic Acids Res. 12, 5355-5367, 1984
A>Title: The DNA sequence of the gene and genetic control sites for the excreted B. subtilis
A:Reference number: A93526; MUID:64272222
A:Accession: A22914
A:Molecule type: DNA
A:Residues: 1-203,'L',205-242 <MUR>
A:Cross-references: EMBL:X00754; NID:g39818; PIDN:CAA25328.1; PID:g685236
A:Experimental source: strain C120
A>Note: the authors translated the codon CAA for residue 29 as Lys and CCA for residue 8
R:Tezuka, H.; Yuuki, T.; Yabuuchi, S.
Agric. Biol. Chem. 53, 2335-2339, 1989
A>Title: Construction of a beta-glucanase hyperproducing *Bacillus subtilis* using the cloned
A:Reference number: A90026
A:Accession: A90026
A:Molecule type: DNA
A:Residues: 1-23,'S',25-82,'S',84-242 <TEZ>
A:Cross-references: DBJ:D00518; NID:g216243; PIDN:BA00405.1; PID:g216244
A:Experimental source: strain Y-25, clone pLE100
R:Yuuki, T.; Tezuka, H.; Yabuuchi, S.
Agric. Biol. Chem. 53, 2341-2346, 1989
A>Title: Purification and some properties of two enzymes from a beta-glucanase hyperproducing
A:Reference number: A90027
A:Contents: annotation
A>Note: source was hyperproducing strain HL-25 with gene from strain Y-25
A:Note: the amino ends of the mature forms of E-1 and E-2 are pyroglutamic acid and glutamic
C:Genetics:
A:Gene: bgls
C:Function:
A>Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenin and carmelum
C:Superfamily: licheninase
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation; F
E-1-28/Domain: signal sequence #status predicted <SIG>
F-29-242/Product: licheninase #status predicted <MAT>
F-29/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) (partial) #status predicted
Query Match 18.4%; Score 245.5; DB 1; Length 242;
Best Local Similarity 35.8%; Pred. No. 1.6e-13;
Matches 59; Conservative 24; Mismatches 71; Indels 11; Gaps 6;
OY 7 FSAGELVTLIEVOYGKFPARKMAASGTSSMFLYONGSETADGRPWVEVDIELGKNP 66
| | | : : : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 87 FDCCGNRSVQYGYGLIYVRKKPAKNTGIYSSFFTYTPT---DSTPWDETDIEFLGDOT 143
| | | : : : | | | | : | | | | : | | | | : | | | | : | | | | :
OY 67 GSFSQNIITGKAGAKDTSEKHNAHVSPPAADQAFHTYGLLEWTPENYVMWTVDGGQEVKRTGEG 126
| | | : : : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 144 TKVGFNYTTNAG--NHKILVDLGFDANAYHTYAFAFMQNSIKIWYDQG-LKMTATNQ 199
| | | : : : | | | | : | | | | : | | | | : | | | | : | | | | :
OY 127 VSNLTGTGCLRFNLMSSEA-AWVGQPEDESKLPLEFOFINWKVKYK 170
| | | : : : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 200 IPTTPGK--IMNNLMNGTGVDWMIGSYNGVN-PLVAHYDMWRKYTK 241
| | | : : : | | | | : | | | | : | | | | : | | | | : | | | | :
RESULT 6
A36910
xylanase, beta(1,3-1,4)-glucanase - Ruminooccus flavefaciens
C:Species: Ruminooccus flavefaciens
C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A36910
R:Plint, H.J.; Martlin, J.; McPherson, C.A.; Daniel, A.S.; Zhang, J.X.
J. Bacteriol. 175, 2943-2951, 1993
A>Title: A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-glucanase domains
A:Reference number: A36910; MUID:93259938
A:Accession: A36910
A:Molecule type: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <PLI>
A:Cross-references: GB:S61204; NID:g385910; PIDN:AAB26620.1; PID:g385911

A: Note: sequence extracted from NCBI backbone (NCBI:U31871, NCBI:U31872)
C: Superfamily: endo-1,4-beta-xylanase homology; Thermotoga xylanase A amino-terminal
F: 4-239/Domain: endo-1,4-beta-xylanase homology <YKL>
F: 259-401/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

```
Query Match          18.2%  Score 242.5;  DB: 2;  Length 802;
Best Local Similarity 34.9%  Pred. No. 1.3e-12;
Matches 58;  Conservative 21;  Mismatches 74;  Indels 13;  Gaps 6;

QY  7  FSGAELTYLEVOYGKFEARKKMAAAGTGVSMFLYONGSEIADGRPWEVDIEVL 66
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db   634  YSGGEFRNNRHHYHYGYSCSMQAMKNDGVSSFFLYTGP---DDNPDEIDIELGKNT 690

QY  67  GSFOSNIITGKAGAKOTSEKHHAVSPAADOAFHTYGLTEWTPNRYRWITVDGGEVRKTEGGQ 126
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db   691  TVQGVNNTYTGQGRK---EKLTYDGESEAYHTFGDMQNPYIAWYADGREYRA---- 743

QY  127  VSNLTGTG-LRFNLMSESA-AWVGQFDESKLPFGFINWYKYK 170
      ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db   744  TQDIPKTPGKIMMANMPEGLTVDWMLKAFN-GRTPHTAHYQWYLYNK 788

RESULT  7
A29091
Licheninase (EC 3.2.1.73) beta - Bacillus amylioliquefaciens
N:Alternate names: 1.3-1,4-beta-D-glucan 4-glucanohydrolase; beta-glucanase; Lichenas
C:Species: Bacillus amylioliquefaciens
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A29091
R:Hotemeister, J.; Kurtz, A.; Bottiss, R.; Knowles, J.
Gene 49, 177-187, 1986
A:Title: The beta-glucanase gene from Bacillus amylioliquefaciens shows extensive homo
A:Reference number: A91564; MUID:87192007
A:Accession: A29091
A:Molecule type: DNA
A:Residues: 1-239 <HOF>
A:Cross-references: GB:M15674; NID:g143009; PIDN:AAA87323.1; PID:g143010
A:Experimental source: strain BE20/78
C:Genetics:
A:Gene: bglA
C:Superfamily: Licheninase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match          17.9%  Score 238.5;  DB: 1;  Length 239;
Best Local Similarity 34.3%  Pred. No. 6.2e-13;
Matches 58;  Conservative 26;  Mismatches 74;  Indels 11;  Gaps 6;

QY  3  SAKDFSGAELTYLEVOYGKFEARKKMAAAGTGVSMFLYONGSEIADGRPWEVDIEVL 62
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db   80  SYNKFDGCENSVQYGYGLYEVRRMKPAKNGIYSSFFLYTGPTE---GTPWDEIDIEFL 136

QY  63  GKNPESFOSNIITGKAGAKOTSEKHHAVSPAADOAFHTYGLTEWTPNRYRWITVDGGEVAKT 122
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db   137  GKDTTKVOGFNNTYTGAG---NHEKFDADGFPDAANAYHTYAADWQPNISIKWYVDGQ-LKHT 192

QY  123  EGGGVSNLTGTGQGLRFNLMSESA-AWVGQFDESKLPFGFINWYKYK 170
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db   193  ATTQIPAPGK--IMANLMNGTGYDDWDLGSTNGVN-PIYATDWMRYK 238

RESULT  8
S23498
Licheninase (EC 3.2.1.73) lichen precursor - Clostridium thermocellum
N:Alternate names: beta-1,3-1,4-glucanase lichen; Lichenase lichen
C:Species: Clostridium thermocellum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S23498; S22137
R:Schlimming, S.; Schwarz, W.H.; Staudenbauer, W.L.
Eur. J. Biochem. 204, 13-19, 1992
A:Title: Structure of the Clostridium thermocellum gene lichen and the encoded beta-1,3
l cellulas.
```

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-238 <GOS>
```



```

Db      87  MTLAKRRDNPGLKSNFYIMYKLEVLKLAANGTGVSSFYLDSDLD-----EIDIE 138
QY      61  VLGRNPGSFGQNIITGKAGAKTSEKHHAVPADQAFHTYGLEWTPYVVRWTVDGOEVR 120
Db     139  WVGDNTOFOGNSFTSKGDTTYDGRGEFHGVDPIDK-FHNTLDNMDKTTWYLDGESVR 197
QY     121  KTEGGQVSNLTGTGQ-----LRFNLWSS--ESAA---WV-GQFDESKLPLQFTIW 165
Db     198  -----VLSNTSSEGYPOSPMYLMMGIWAGDPDPAAGTIEMAGGETNYNDAPFTMYIEK 251
QY     166  VKVYKYPGQEGSGDFTLDWTFDNDFTDGSRMKGD 202
Db     252  VIVDYSTGKKYTYGDSGSW-ESIEADGSGIYGRYD 287

```

RESULT 15

```

S34804
exok protein - Rhizobium meliloti megaplasmid 2
C:Species: Rhizobium meliloti
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: S34804; S30072
R:Becker, A.; Kleickmann, A.; Arnold, W.; Puhler, A.
Mol. Gen. Genet. 238: 145-154, 1993
A:Title: Analysis of the Rhizobium meliloti exoH/exoK/exoL fragment: ExoK shows homology
A:Reference number: S34803; MID:93241147
A:Accession: S34804
A:Molecule type: DNA
A:Residues: 1-268 <BEC>
A:Cross-references: EMBL:Z17219
A:Note: the authors translated the codon GTA for residue 208 as Leu
C:Genetics:
A:Gene: exok
A:Genome: plasmid
C:Superfamily: licheninase

```

```

Query Match      13.5%; Score 180; DB 2; Length 268;
Best Local Similarity 33.8%; Pred. No. 6.3e-08;
Matches 46; Conservative 26; Mismatches 52; Indels 12; Gaps 6;

```

```

QY      20  YGKFEARMKMAASGVSSMFLVONGSEIADGRPWVEVDIEVLGKNPQSFQSN-IITGKA 78
Db     104  YGYEARIKADSGSLNSAFETYIGP--ADKKPHDEIDFEVLGKNYAKVOINQYVASAKG 160
QY      79  GAOKTSEKHHAVPADQAFHTYGLEWTPYVVRWTVDGOEVRK-TEGGQVSNLTGTGLR 137
Db     161  GNEFLAD---VFGAGNGFNDAFVWEKNRIRIYVNGELVHEVTDPAKIP--VNAOKIF 214
QY     138  FNLMSSESAA-WVGQF 152
Db     215  FSLMGTDITLTDMMGTG 230

```

Search completed: August 23, 2001, 11:15:07
 Job time: 94 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:14:33 ; Search time 32.82 Seconds
(without alignments)
575.603 Million cell updates/sec

Title: US-09-654-652A-1

Perfect score: 248
Sequence: 1 MSAKDFSGAELTYLLEVOY.....TRKGESFNQVPRDEPPAP 248

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	99.6	349	2 A44507	licheninase (EC 3.1.1.1)
2	8	3.2	381	2 S29560	fructose-bisphosphatase
3	8	3.2	407	2 T06408	probable fructose-1,6-bisphosphatase
4	8	3.2	429	2 T07815	S-locus-specific protein
5	8	3.2	429	2 D70784	probable Rieske [2-iron-2-sulfur] ferredoxin
6	7	2.8	95	2 S55439	hypothetical protein
7	7	2.8	115	2 F84469	probable glycine-1-aminotransferase
8	7	2.8	119	2 S64079	probable membrane protein
9	7	2.8	304	2 F69864	hypothetical protein
10	7	2.8	317	2 F82672	ATP sulfurylase, cytoplasmic
11	7	2.8	348	2 T08814	1,3-beta-glucanase
12	7	2.8	396	2 S15992	flavohemoglobin hmoA
13	7	2.8	396	2 F85900	hypothetical protein
14	7	2.8	420	2 T36193	probable salicylate synthase
15	7	2.8	440	2 D64186	conserved hypothetical protein
16	7	2.8	468	1 TVMSE2	transcription factor
17	7	2.8	469	1 TVHUE2	transcription factor
18	7	2.8	474	2 S75464	hypothetical protein
19	7	2.8	546	2 B32688	beta-galactosidase
20	7	2.8	611	2 F70325	conserved hypothetical protein
21	7	2.8	637	2 E70867	probable fadD32 protein
22	7	2.8	677	2 A32611	beta-galactosidase
23	7	2.8	715	2 H84799	hypothetical protein
24	7	2.8	955	2 T48515	hypothetical protein
25	7	2.8	994	1 J00151	myosin heavy chain
26	7	2.8	1026	2 A48995	exocystalline su
27	7	2.8	1171	2 F83110	hypothetical protein
28	7	2.8	1495	2 T48429	hypothetical protein
29	6	2.4	58	2 E81202	hypothetical protein

30	6	2.4	60	2 G81777	probable periplasmic protein
31	6	2.4	66	2 I40095	hypothetical protein
32	6	2.4	69	2 H81078	4-oxalocrotonate t
33	6	2.4	77	1 F69115	hypothetical protein
34	6	2.4	80	2 S49552	hypothetical protein
35	6	2.4	88	2 T06988	pathogen-induced p
36	6	2.4	90	2 B83542	hypothetical protein
37	6	2.4	91	2 T31202	hypothetical protein
38	6	2.4	94	2 T35283	hypothetical protein
39	6	2.4	95	2 S46537	pathogen-inducible
40	6	2.4	96	2 G33730	Ig kappa chain v r
41	6	2.4	96	2 A81031	hypothetical protein
42	6	2.4	103	2 S19975	Ig kappa chain v r
43	6	2.4	106	2 E72239	hypothetical protein
44	6	2.4	109	2 T34639	hypothetical protein
45	6	2.4	111	2 G30502	Ig kappa chain v r

ALIGNMENTS

RESULT 1
A44507
Licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Teather, R.M.; Erfle, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1,3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-references: EMBL:M3676; NID:g148575; PIDN:AAA24896.1; PID:g148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 99.6%; Score 247; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.9e-250; Indels 0; Gaps 0;
Matches 247; Conservative 0; Mismatches 0;
QY 2 VSANKDFSGAELTYLLEVOYGFKFAKMAAGTYSMFLYONGSEIADGRPWVEDIEV 61
DB 25 VSANKDFSGAELTYLLEVOYGFKFAKMAAGTYSMFLYONGSEIADGRPWVEDIEV 84
QY 62 LCKNPGSQSNITITGKAGAKTSEKHHAVSPADQAFHTYGLEWTPNNYRWTVDQGEYRK 121
DB 85 LCKNPGSQSNITITGKAGAKTSEKHHAVSPADQAFHTYGLEWTPNNYRWTVDQGEYRK 144
QY 122 TEGGVSNLTGTGGLRFNIMSSSAWVGQFDESKLPFLQFINWVKYKYYTGTGGEFGSD 181
DB 145 TEGGVSNLTGTGGLRFNIMSSSAWVGQFDESKLPFLQFINWVKYKYYTGTGGEFGSD 204
QY 182 FTLDWTDNFDFTDGSRWKGDWTFPGNRYDLTDKNYSRDGMLIALPRKGESFNQVPR 241
DB 205 FTLDWTDNFDFTDGSRWKGDWTFPGNRYDLTDKNYSRDGMLIALPRKGESFNQVPR 264
QY 242 RDDEPPAP 248
DB 265 RDDEPPAP 271
RESULT 2
fructose-bisphosphatase (EC 3.1.3.11) - garden pea (fragment)
C:Species: Pisum sativum (garden pea)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S29560
R:Carriaco, J.L.; Chueca, A.; Hermoso, R.; Lazaro, J.J.; Ramos, J.L.; Sahrawy, M.; Pr
submitted to the EMBL Data Library, October 1992
A:Description: Cloning, structure and expression of a pea cDNA clone coding for a pho

A:Reference number: S29560
 A:Accession: S29560
 A:Molecule type: mRNA
 A:Residues: 1381 <CAR>
 A:Cross-references: EMBL:X68826; NID:g20716; PIDN:CAA48719.1; PID:g20717
 C:Superfamily: fructose-bisphosphatase
 C:Keywords: phosphoric monoester hydrolase

Query Match 3.2%; Score 8; DB 2; Length 361;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 SNLTGTGG 135
 |||||

DB 86 SNLTGTGG 93

RESULT 3
 T06408
 probable fructose-bisphosphatase (EC 3.1.3.11) precursor - garden pea chloroplast
 C:Species: chloroplast Pisum sativum (garden pea)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Jun-1999
 C:Accession: T06408
 R:Hahn, T.R.; Dong, S.M.; Rhim, J.H.
 submitted to the EMBL Data Library, February 1999
 A:Description: cDNA sequence and red light modulated expression of chloroplast fructose-
 A:Reference number: Z15655
 A:Accession: T06408
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-407 <HAH>
 A:Cross-references: EMBL:L34806; NID:g609560; PIDN:AAD10213.1; PID:g609561
 A:Experimental source: cv. Giant
 C:Genetics:
 A:Genome: chloroplast
 C:Superfamily: fructose-bisphosphatase
 C:Keywords: chloroplast; phosphoric monoester hydrolase
 F:1-50/Domain1: transit peptide (chloroplast) #status predicted <TNP>
 F:51-407/Product: fructose-bisphosphatase #status predicted <MNT>

Query Match 3.2%; Score 8; DB 2; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 SNLTGTGG 135
 |||||

DB 112 SNLTGTGG 119

RESULT 4
 T07815
 S-locus-specific glycoprotein S8 - radish (fragment)
 C:Species: Raphanus sativus (radish)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
 C:Accession: T07815
 R:Sakamoto, K.; Kusaba, M.; Nishio, T.
 Mol. Gen. Genet. 258, 397-403, 1998
 A:Title: Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gen
 A:Reference number: Z16146; MUID:98311079
 A:Accession: T07815
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-429 <SAK>
 A:Cross-references: EMBL:AB009683; NID:g3327851; PIDN:BAA31730.1; PID:g3327852
 C:Genetics:
 A:Gene: SLG(S8)
 C:Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology
 C:Keywords: glycoprotein
 F:32-427/Domain: S-locus-specific glycoprotein homology <SSG>

Query Match 3.2%; Score 8; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 TDXNIYSR 220
 |||||

DB 253 TDXNIYSR 260

RESULT 5
 D70784
 probable Rieske [2Fe-2S] iron-sulfur protein cyob - Mycobacterium tuberculosis (strai
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: D70784
 R:Coile, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98255987
 A:Accession: D70784
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-429 <COL>
 A:Cross-references: GB:Z70283; GB:AL123456; NID:g3261561; PIDN:CAA94264.1; PID:e23357
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: qcrA
 C:Superfamily: Rieske [2Fe-2S] homology
 C:Keywords: 2Fe-2S; metalloprotein; Rieske iron-sulfur protein
 F:343-389/Domain1: Rieske [2Fe-2S] homology <RSK>
 F:353,355,372,375/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #statu

Query Match 3.2%; Score 8; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RKQGESFN 237
 |||||

DB 335 RKQGESFN 342

RESULT 6
 S55439
 hypothetical protein ykf - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
 C:Accession: S55439; E70061
 R:Glaser, P.; Danchin, A.
 submitted to the EMBL Data Library, May 1995
 A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro
 A:Reference number: S55414
 A:Accession: S55439
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-95 <GIA>
 A:Cross-references: EMBL:Z49782; NID:g853752; PIDN:CAA89886.1; PID:g853778
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 A.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Enllich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koeltter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rooba, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akuchul, M.; Yamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E70061
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-95 <KUN>
 A:Cross-references: GB:299122; GB:AL009126; NID:92636029; PIDN:CAB15716.1; PID:e1184605;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ywxf

Query Match 2.8%; Score 7; DB 2; Length 95;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKMMAA 32
 DB 71 RKMMAA 77

RESULT 7

Probable glycine-rich protein [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84469
 R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.;
 Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-766, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; MUID:20083487
 A:Accession: F84469
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-115 <STO>
 A:Cross-references: GB:AE002093; NID:94581171; PIDN:AAD24654.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g05530
 A:Map position: 2

Query Match 2.8%; Score 7; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 AAASGV 36
 DB 22 AAASGV 28

RESULT 8

Probable membrane protein YGL072c - yeast (*Saccharomyces cerevisiae*)
 S64079
 N:Alternate names: hypothetical protein G3263
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jun-2000
 C:Accession: S64079
 R:Rieger, M.; Mueller-Auer, S.; Bruckner, M.; Schaefer, M.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64071
 A:Accession: S64079
 A:Molecule type: DNA
 A:Residues: 1-119 <RIB>
 A:Cross-references: EMBL:Z72596; NID:g1322584; PID:e243291; PID:g1322587; GSPDB:GN00007;
 A:Experimental source: strain 5288C
 C:Genetics:
 A:Gene: MIPS:YGL072C
 A:Map position: 7L
 C:Superfamily: *Saccharomyces cerevisiae* probable membrane protein YGL072c
 C:Keywords: transmembrane protein

F:35-51/Domain: transmembrane #status predicted <TM>

Query Match 2.8%; Score 7; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 FSGAEY 13
 DB 98 FSGAEY 104

RESULT 9

hypothetical protein YKtd - *Bacillus subtilis*
 F69864
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: F69864
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: F69864
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-304 <KUN>
 A:Cross-references: GB:299111; GB:AL009126; NID:92633699; PIDN:CAB13342.1; PID:e11850
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yktd

Query Match 2.8%; Score 7; DB 2; Length 304;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 KMAASG 34
 DB 240 KMAASG 246

RESULT 10

APP sulfurylase, small subunit XF1500 [imported] - *Xylella fastidiosa* (strain 9a5c)
 F82672
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: F82672
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20965717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: F82672
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-317 <SIM>
 A:Cross-references: GB:AE003980; GB:AE003849; NID:g9106531; PIDN:AAF84309.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
 as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Jungheiler, M.L.; Kemp, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurahoe, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savaasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., M.A.; da Silveira
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1500
C:Superfamily: modulation protein nodB

Query Match 2.8%; Score 7; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 220 RDGMIL 226
|||
Db 234 RDGMIL 240

RESULT 11
T08814

1.3-beta-glucanase (EC 3.2.1.-) SGNI - soybean

C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999

C:Accession: T08814

R:Cheong, Y.H.; Kim, C.Y.; Chun, H.J.; Hong, J.C.; Bahk, J.D.; Cho, M.J.
submitted to the EMBL Data Library, November 1995

A:Description: Isolation and characterization of the soybean beta-1,3-glucanase induced

A:Reference number: Z16477

A:Accession: T08814

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-348 <CHE>

A:Cross-references: EMBL:U01323; NID:G1403674; PID:G1403675

A:Experimental source: strain Williams

C:Genetics:

A:Gene: SGNI

A:Introns: 32/1

A>Note: induced by wounding and fungal elicitor

C:Superfamily: beta-1,3-glucanase

C:Keywords: glycosidase; hydrolase

Query Match 2.8%; Score 7; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 TGRGAQ 81
|||||
Db 28 TGRGAQ 34

RESULT 12

S15992
flavohemoglobin hmp - Escherichia coli

N:Alternate names: ferrisiderophore reductase fsrB; flavohemoprotein

C:Species: Escherichia coli

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 03-Mar-2000

C:Accession: S15992; S21161; S25241; G65032

R:Vasudevan, S.G.; Armarego, W.L.F.; Shaw, D.C.; Lilley, P.E.; Dixon, N.E.; Poole, R.K.
Mol. Gen. Genet. 226, 49-58, 1991

A:Title: Isolation and nucleotide sequence of the hmp gene that encodes a haemoglobin-11

A:Reference number: S15991; MUID:91238719
A:Accession: S15992
A:Molecule type: DNA
A:Residues: 1-396 <YAS>
A:Cross-references: EMBL:X58872; NID:941730; PID:CAA1682.1; PID:941731

R:Andrews, S.C.; Shipley, D.; Keen, J.N.; Findlay, J.B.C.; Harrison, P.M.; Guest, J.R
FEBS Lett. 302, 247-252, 1992

A:Title: The haemoglobin-like protein (HMP) of Escherichia coli has ferrisiderophore

A:Reference number: S21161; MUID:92290008

A:Accession: S21161

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-37, 'Q', 39-64; 359-396 <AND>

R:Pliemann, M.D.; Stauffer, G.V.

Gene 22, 9-18, 1983

A:Title: Characterization of the Escherichia coli gene for serine hydroxymethyltransf

A:Reference number: S25241; MUID:83235562

A:Accession: S25241

A:Molecule type: DNA

A:Residues: 1-10 <PLA>

A:Cross-references: EMBL:X58872

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G65032

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-396 <BLAT>

A:Cross-references: GB:AE000341; GB:U00096; NID:G1788899; PID:AACT5605.1; PID:G17889

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: hmpA

C:Superfamily: flavohemoglobin; cytochrome-b5 reductase homology; globin homology

C:Keywords: chromoprotein; flavoprotein; heme; iron; metalloprotein; monomer

F:2-136/Domain: globin homology <GLB>

F:157-390/Domain: cytochrome-b5 reductase homology <CBR>

F:53/Binding site: oxygen (Gln) (distal axial ligand) #status predicted

F:85/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 2.8%; Score 7; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 EGGQVSN 129
|||||
Db 225 EGGQVSN 231

RESULT 13

F85900
hypothetical protein hmpA [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: F85900

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grobeck, E.J.; Davis, N.W.; Llim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F85900

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-396 <STO>

A:Cross-references: GB:AE005174; NID:G12516962; PID:AA657666.1; GSPDB:GN00145; UNCP:

C:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: hmpA

C:Superfamily: flavohemoglobin; cytochrome-b5 reductase homology; globin homology

Query Match 2.8%; Score 7; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 EGGQVSN 129

|||||||
DB 225 EGGVSN 231

RESULT 14
T36193
probable salicylate hydroxylase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T36193
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21600
A:Accession: T36193
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-420 <SAU>
A:Cross-references: EMBL:AL055707; PIDN:CAB38889.1; GSPDB:GN00070; SCODEB:SCE29.14C
C:Genetics:
A:Gene: SCODEB:SCE29.14C
C:Superfamily: tetracycline 6-hydroxylase

Query Match 2.8%; Score 7; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 226 LALTRKG 232
|||||||
DB 20 LALTRKG 26

RESULT 15
D64186
conserved hypothetical protein H1154 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: D64186
R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Snall, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: D64186
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-440 <TIGR>
A:Cross-references: GB:U032795; GB:I42023; NID:g1574708; PIDN:AAC22809.1; PID:g1574711; T
C:Superfamily: Bacillus subtilis sodium-glutamate symporter homolog yhcL

Query Match 2.8%; Score 7; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GOVSMLT 131
|||||||
DB 150 GOVSMLT 156

Search completed: August 23, 2001, 11:18:02
Job time: 209 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:15:13 ; Search time 19.47 Seconds
(without alignments)
436.331 Million cell updates/sec

Title: US-09-654-652a-1
Perfect score: 248
Sequence: 1 MVSARDFSGAELTYLEVOY.....TRKGESFNGQVPRDEPAP 248

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size: 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	99.6	349	1 GUB_FIBSU	P17989 fibroblacter
2	8	3.2	317	1 CYSD_RHIME	P56892 rhizobium m
3	8	3.2	407	1 F16P_PEA	P46275 pisum sativ
4	8	3.2	429	1 OGRA_MYCTU	Q10387 mycobacteri
5	7	2.8	95	1 YWKF_BACSU	P5874 bacillus su
6	7	2.8	119	1 YGH2_YEAST	P3161 saccharomyc
7	7	2.8	301	1 NODP_RHITS	P72338 rhizobium s
8	7	2.8	396	1 HMPA_ECOLI	P24232 escherichia
9	7	2.8	440	1 YDJN_HABIN	P45079 haemophilus
10	7	2.8	468	1 ETS2_MOUSE	P15037 mus musculi
11	7	2.8	469	1 ETS2_HUMAN	P15036 homo sapien
12	7	2.8	546	1 BGAM_HUMAN	P16279 homo sapien
13	7	2.8	677	1 BGAL_HUMAN	P22467 dictyosteli
14	7	2.8	994	1 MISA_DICDI	P41368 staphylococ
15	7	2.8	1024	1 SYIP_STAAU	P35828 caulobacter
16	7	2.8	1025	1 SIAP_CAOCR	O01482 tritium ae
17	6	2.4	88	1 WIRA_MHEAT	O080294 bacterioph
18	6	2.4	95	1 WHED_BPTEL	P41370 staphylococ
19	6	2.4	131	1 YTL2_STAAU	O69529 mycobacteri
20	6	2.4	132	1 FOLB_MYCLE	P39258 bacterioph
21	6	2.4	133	1 Y05C_BPT4	P49100 oryza sativ
22	6	2.4	137	1 CYB5_ORYSA	P01034 homo sapien
23	6	2.4	146	1 CYTC_HUMAN	P01092 macaca mula
24	6	2.4	146	1 CYTC_MACMU	O51726 methanococ
25	6	2.4	148	1 CYTC_BOVIN	P42245 bacillus su
26	6	2.4	157	1 FKBL_MERYA	P28989 equine herp
27	6	2.4	158	1 YCBM_BACSU	Q10613 mycobacteri
28	6	2.4	160	1 VG34_HSVB	P40946 drosophila
29	6	2.4	161	1 YC87_MYCTU	P26990 gallus galli
30	6	2.4	174	1 ARF3_DROME	P26438 homo sapien
31	6	2.4	174	1 ARF6_CHICK	P51645 xenopus lae
32	6	2.4	174	1 ARF6_HUMAN	
33	6	2.4	174	1 ARF6_XENILA	

34	6	2.4	176	1 VEGP_PIG	P53715 sus scrofa
35	6	2.4	178	1 MYCB_RAT	P15063 rattus norv
36	6	2.4	179	1 UBCH_XENILA	P56616 xenopus lae
37	6	2.4	191	1 TERE_ALCSP	P18782 alciigenes
38	6	2.4	193	1 FLHC_PROMI	O34202 proteus mir
39	6	2.4	194	1 FLHC_XENNE	O9x9f3 xenorhabdus
40	6	2.4	199	1 PEXH_YEAST	P40155 saccharomyc
41	6	2.4	202	1 Y105_MYCPN	P75528 mycoplasma
42	6	2.4	204	1 PRTB_SCYLI	P15369 scytalidium
43	6	2.4	206	1 YPTB_YEAST	P01123 saccharomyc
44	6	2.4	209	1 RR3_CRATE	P16631 gracillaria
45	6	2.4	215	1 US11_HCMVA	P09727 human cytom

ALIGNMENTS

RESULT 1	GUB_FIBSU	STANDARD:	PRT:	349 AA.
AC	P17989:			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)			
DE	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).			
OS	Fibroblacter succinogenes (Bacteroides succinogenes).			
OC	Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;			
OC	Fibrobacter.			
OX	NCBI_TaxID=833;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.			
RC	STRAIN-ISOLATE S85:			
RK	MEDLINE-90299807; PubMed-2193918;			
RA	Teather R.M., Erile J.D.;			
RT	"DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";			
RL	J. Bacteriol. 172:3837-3841(1990).			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M33676; AAA24896.1; -.			
DR	PIR: A44507; A44507.			
DR	HSSP; P23904; IAKK.			
DR	InterPro: IPR000757; -.			
DR	Pfam: PF00722; Glyco_hydro.16; 1.			
DR	PROSITE: PS01034; GLYCOSYL_HYDROL.16; 1.			
KW	Hydrolase; Glycosidase; Signal; Repeat.			
FT	SIGNAL	1	27	
FT	CHAIN	28	349	
FT	ACT_SITE	79	79	BETA-GLUCANASE.
FT	ACT_SITE	83	83	NUCLEOPHILE (BY SIMILARITY).
FT	DOMAIN	271	307	PROTON DONOR (BY SIMILARITY).
FT	REPEAT	271	277	5 X 7 AA TANDEM REPEATS OF P-X-S-S-S-S-X.
FT	REPEAT	278	284	
FT	REPEAT	285	291	
FT	REPEAT	292	298	
FT	REPEAT	301	307	
SO	SEQUENCE	349 AA;	37737 MW;	16DCAF5BDEFC578A CRC64;

Query Match 99.6%; Score 247; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 2e-250;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VSADFGAELTYLEEVQYGFARMKMAAAGTSSMFLYQNGSEIADGPRWVEIDV 61
| | | | |
DB 25 VSADFGAELTYLEEVQYGFARMKMAAAGTSSMFLYQNGSEIADGPRWVEIDV 84
| | | | |
QY 62 LKNGPSFOSNIITGKAGAKTSEKHHAVSPADQAFHTYGLWTPVYVNTVDGOEVRK 121
| | | | |
DB 85 LKNGPSFOSNIITGKAGAKTSEKHHAVSPADQAFHTYGLWTPVYVNTVDGOEVRK 144
| | | | |
QY 122 TEGGVSNLTGTOGLRRNLMSSSAAMVGOFPDESKLPFOFINNVKYYKTPGCGEGSD 181
| | | | |
DB 145 TEGGVSNLTGTOGLRRNLMSSSAAMVGOFPDESKLPFOFINNVKYYKTPGCGEGSD 204
| | | | |
QY 182 FTLDWNTDNFTFGDSRNGKGDWTFDGNRVLDTDKNITSRDCMLTLATLRKGOSEFNGOVP 241
| | | | |
DB 205 FTLDWNTDNFTFGDSRNGKGDWTFDGNRVLDTDKNITSRDCMLTLATLRKGOSEFNGOVP 264
| | | | |
QY 242 RDDEPAP 248
| | | | |
DB 265 RDDEPAP 271
| | | | |

RESULT 2
CYSD_RHIME STANDARD; PRT; 317 AA.
ID CYSD_RHIME STANDARD; PRT; 317 AA.
AC P56892;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SULFATE ADENYLYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLYLTRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT).
GN Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Rhizobiaceae; Sinorhizobium.
RN NCB1_TaxID:382;
RP SEQUENCE FROM N.A.
RC STRAIN-1021;
RX MEDLINE-99395034; PubMed-10464198;
RA Adola A.P., Willis M.G., Wang R.C., Long S.R.;
RT "Reduction of adenosine-5'-phosphosulfate instead of 3'-phosphoadenosine-5'-phosphosulfate in cysteine biosynthesis by Rhizobium meliloti and other members of the family Rhizobiaceae";
RT J. Bacteriol. 181:5280-5287(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-1021;
RA Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA Boistard P., Guzy J., Kahn D., Thebaud P., Goffeau A.,
RA Purnelle B., Pohl T., Boche G., Schneider S., Portetelle D.,
RA Vandenbol M., Puehler A., Becker A., Weidner S.;
RL Submitted (May-2000) to the SWISS-PROT data bank.
CC -1- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE + ADENYLYLSULFATE.
CC -1- PATHWAY: FIRST STEP IN THE SULFATE ACTIVATION PATHWAY. THIS REACTION OCCURS EARLY IN THE REDUCTIVE BRANCH OF THE CYSTEINE BIOSYNTHETIC PATHWAY.
CC -1- SUBUNIT: HETERODIMER COMPOSED OF CYSD, THE SMALLER SUBUNIT, AND CYSDN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: AF158023; AAD55760.1; -
CC DR InterPro: IPR002500; -
CC

DR Pfam: PF01507; PAPS_reduct; 1.
KW Cysteine biosynthesis; Transferase; Nucleotidyltransferase.
FT CONFLICT 260 260 R -> A (IN REF. 1).
SQ SEQUENCE 317 AA: 36429 MW: 093f2724d22841CB CRC64;

Query Match 3.2%; Score 8; DB 1; Length 317;
Best local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RDGMLIA 227
| | | | |
DB 234 RDGMLIA 241
| | | | |

RESULT 3
F16P_PEA STANDARD; PRT; 407 AA.
ID F16P_PEA STANDARD; PRT; 407 AA.
AC P46275; Q37263;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FRUCTOSE-1,6-BISPHOSPHATASE, CHLOROPLAST PRECURSOR (EC 3.1.3.11)
DE (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPASE).
GN FBP.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Pisum.
OX NCB1_TaxID:3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. GIANT; TISSUE-leaf;
RX MEDLINE-95175626; PubMed-7870839;
RA Dong S.M., Rhim J.H., Hahn T.R.;
RT "Nucleotide sequence analysis of a cDNA encoding chloroplastic fructose-1,6-bisphosphatase from pea (Pisum sativum L.).";
RL Plant Physiol. 107:313-314(1995).
RN [2]
RP SEQUENCE OF 27-407 FROM N.A.
RC STRAIN-CV. LINCOLN; TISSUE-leaf;
RX MEDLINE-94297517; PubMed-7764999;
RA Carrasco J.L., Cheuca A., Prado F.E., Hermoso R., Lazaro J.J.,
RA Ramos J.L., Sahravy M., Lopez Gorge J.;
RT "Cloning, structure and expression of a pea cDNA clone coding for a photosynthetic fructose-1,6-bisphosphatase with some features different from those of the leaf chloroplast enzyme";
RT Planta 193:494-501(1994).
RL [1]
CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 1,6-BISPHOSPHATE + H(2)O -> D-FRUCTOSE 6-PHOSPHATE + ORTHOPHOSPHATE.
CC -1- PATHWAY: THE CHLOROPLAST ISOZYME TAKES PART IN THE REGENERATION OF RIBULOSE BISPHOSPHATE IN THE PHOTOSYNTHETIC CARBON REDUCTION CYCLE (CALVIN CYCLE).
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
CC -1- INDUCTION: LIGHT ACTIVATION THROUGH PH CHANGES, MG(2+) LEVELS AND ALSO BY LIGHT-MODULATED REDUCTION OF ESSENTIAL DISULFIDE GROUPS VIA THE FERREDOXIN-THIOREDOXIN F SYSTEM (BY SIMILARITY).
CC -1- MISCELLANEOUS: IN PLANTS THERE ARE TWO FBPAE ISOZYMES: ONE IN THE CYTOSOL AND THE OTHER IN THE CHLOROPLAST.
CC -1- SIMILARITY: BELONGS TO THE FBPAE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: L34806; AAD10213.1; -
CC DR EMBL: X68826; CAA48719.1; -
CC DR HSSP: P22418; ISPI.
CC

DR InterPro: IPR000146; -
DR Pfam: PF00316; FBpase; 1.
DR PRINTS: PR00115; FBPHPTASE.
DR PRINTS: PR00377; INBPHPTASE.
DR PROSITE: PS00124; FBpase; 1.
KW Hydrolase; Carbohydrate metabolism; Multigene family; Chloroplast;
FT Transit peptide; Calvin cycle.
FT TRANSIT 1 50
FT CHAIN 51 407
FT ACT_SITE 349 349
FT DISULFID 223 228
FT CONFLICT 82 82
FT CONFLICT 160 160
FT CONFLICT 247 247
FT CONFLICT 282 282
SO SEQUENCE 407 AA; 44511 MW; B60E9164F1F6F9D CRC64;

Query Match 3.2%; Score 8; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 128 SMLGTGCG 135
Db 112 SMLGTGCG 119

RESULT 4
OCRA_MYCTU STANDARD; PRT; 429 AA.
AC Q10387;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT (RIESKE IRON-SULFUR PROTEIN).
GN OCRA OR RV2195 OR MYCY190.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX. THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF THE UBIQUINOL-CYTOCHROME C COMPLEX ARE: CYTOCHROME B, THE RIESKE PROTEIN AND A CYTOCHROME C SUBUNIT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA, BACTERIAL, CHLOROPLAST).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Z70283; CA94264.1; -
DR Tubercultist; RV2195; -
DR InterPro: IPR001281; -
DR PROSITE: PS00199; RIESKE_1; FALSE_NBS.
DR PROSITE: PS00200; RIESKE_2; 1.
KW Electron transport; Iron-sulfur; Oxidoreductase; Transmembrane.
FT TRANSMEM 96 116
FT TRANSMEM 137 157
FT TRANSMEM 207 227
FT METAL 353 353
FT METAL 355 355
FT METAL 372 372
FT METAL 375 375
FT DISULFID 358 374
SO SEQUENCE 429 AA; 46922 MW; C22C87EC283FE0EE CRC64;

Query Match 3.2%; Score 8; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 230 RKQGESFN 237
Db 335 RKQGESFN 342

RESULT 5
YWK_F_BACSU STANDARD; PRT; 95 AA.
AC P45874;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 10.1 KDA PROTEIN IN PRR-SPOLIR INTERGENIC REGION.
GN YWK_F.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Glaser P., de la Fuente V., Danchin A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 75-95 FROM N.A.
RC STRAIN=168;
RA Glaser P., Danchin A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

QY 26 RMKAAA 32
 ID 111111
 DB 71 RMKAAA 77

RESULT 6
 YGH2_YEAST STANDARD; PRT; 119 AA.
 ID YGH2_YEAST
 AC P53161;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEMICAL 13.4 KDA PROTEIN IN HSF1-AFT1 INTERGENIC REGION.
 GN YGL072C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces;
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=97435481; PubMed=9290212;
 RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
 RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
 chromosome VII.";
 RL Yeast 13:1077-1090(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: 272596; CA96778.1; -;
 DR SGD: S0003040; YGL072C.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 30 50 POTENTIAL.
 SQ SEQUENCE 119 AA; 13388 MW; FID2F040C1C833A6 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 119;
 Best Local Similarity 100.0%; Pred. No. 5.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 FSGAEY 13
 ID 111111
 DB 98 FSGAEY 104

RESULT 7
 NODP_RHIS3 STANDARD; PRT; 301 AA.
 ID NODP_RHIS3
 AC P72338;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SULFATE ADENYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLATE
 DE TRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT) (MODULATION PROTEIN
 DE P).
 GN NODP.
 OS Rhizobium sp. (strain N33).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=103798;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cloutier J.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER
 CC TO MOD FACTOR.
 CC -----

CC -1- CATALYTIC ACTIVITY: ATP + SULFATE = PYROPHOSPHATE +
 CC ADENYLYLSULFATE.
 CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYMS, NODP AND NODQ, MAY BE
 CC PHYSICALLY ASSOCIATED (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U53327; AAB16901.1; -;
 DR InterPro: IPR002500; -;
 DR Pfam: PF01507; PAPS_Reduct; 1.
 KW Nodulation; Transferase; Nucleotidyltransferase.
 SQ SEQUENCE 301 AA; 35008 MW; 3582F7EAB0861431 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 301;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RDGMILL 226
 ID 111111
 DB 218 RDGMILL 224

RESULT 8
 HMPA_ECOLI STANDARD; PRT; 396 AA.
 ID HMPA_ECOLI
 AC P24232;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE FLAVOHEMOPROTEIN (HEMOGLOBIN-LIKE PROTEIN) (FLAVOHEMOGLOBIN)
 DE (DIHYDROPTERIDINE REDUCTASE (EC 1.6.99.7)) (FERRISIDEROPHORE REDUCTASE
 DE B) (NITRIC OXIDE DIOXYGENASE) (NOD).
 GN HMP OR HMPA OR FSRB.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
 RC STRAIN-K12;
 RX MEDLINE=91238719; PubMed=2034230;
 RA Vasudevan S.G., Armarego W.L.F., Shaw D.C., Lilley P.E., Dixon N.E.,
 RA Poole R.K.;
 RT "Isolation and nucleotide sequence of the hmp gene that encodes a
 RT haemoglobin-like protein in Escherichia coli K-12.";
 RL Mol. Gen. Genet. 226:49-58(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97349960; PubMed=9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Saito G., Satoh Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,

RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
RT - K12 genome corresponding to 50,0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE OF 1-10 FROM N.A.
RX MEDLINE=83235562; PubMed=6190704;
RA Plamann M.D., Stauffer G.V.;
RT "Characterization of the *Escherichia coli* gene for serine
RT hydroxymethyltransferase.";
RL Gene 22:9-18(1983).
RN [5]
RP PARTIAL SEQUENCE. AND CHARACTERIZATION.
RC STRAIN-K12;
RX MEDLINE=92290008; PubMed=1601132;
RA Andrews S.C., Shipley D., Keen J.N., Findlay J.B.C., Harrison P.M.,
RA Guest J.R.;
RT "The haemoglobin-like protein (HMP) of *Escherichia coli* has
RT ferrioxido-reductase activity and its C-terminal domain shares
RT homology with ferredoxin NADP+ reductases.";
RL FEBS Lett. 302:247-252(1992).
RN [6]
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [7]
RP SEQUENCE OF 1-20, AND CHARACTERIZATION.
RC MEDLINE=98393652; PubMed=9724711;
RA Gardner P.R., Gardner A.M., Martin L.A., Salzman A.L.;
RT "Nitric oxide dioxygenase: an enzymic function for flavohemoglobin.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10378-10383(1998).
CC -1- FUNCTION: POSSESSES DIHYDROPERIDINE REDUCTASE ACTIVITY. ALSO HAS
CC NITRIC OXIDE DIOXYGENASE ACTIVITY.
CC -1- FUNCTION: SEEMS TO REDUCE THE COMPLEXED FE+3 IRON OF SIDEROPHORES
CC TO FE+2, THUS RELEASING IT FROM THE CHELATOR.
CC -1- CATALYTIC ACTIVITY: NAD(P)H + 6,7-DIHYDROPERIDINE = NAD(P)(+) +
CC 5,6,7,8-TETRAHYDROTERIDINE.
CC -1- DOMAIN: MONOMER.
CC -1- DOMAIN: CONSISTS OF TWO DISTINCT DOMAINS; ONE IS A HEME-CONTAINING
CC OXYGEN BINDING DOMAIN IN THE N-TERMINAL REGION AND THE OTHER IS AN
CC FAD-CONTAINING REDUCTASE DOMAIN FOUND IN THE C-TERMINAL REGION.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY. TWO DOMAINS
CC FLAVOHEMOPROTEINS SUBFAMILY.
CC -1- SIMILARITY: TO A NUMBER OF FAD/NAD(P) FLAVOPROTEIN
CC OXIDOREDUCTASES.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on lists
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56872; CAA41682.1; -
DR EMBL: AE000341; AAC75605.1; -
DR EMBL: D90885; BAA16460.1; -
DR EMBL: J01620; AAA23911.1; -
DR PIR: S15992; S15992.
DR PIR: S21161; S21161.
DR HSSP: P04252; 1YHB.
DR Ecogen; EG10456; Jimp.
DR InterPro: IPR000971; -
DR InterPro: IPR001433; -
DR Pfam: PF00042; globin; 1.
DR Pfam: PF00175; oxidored_fad; 1.
DR PROSITE: PS01033; GLOBIN; 1.
KW Oxidoreductase; NADP; Heme; Flavoprotein; FAD; Iron transport.

```

FT DOMAIN 1 134 GLOIN-
FT METAL 53 53 IROn (HEME DISTAL LIGAND)
FT FT (BY SIMILARITY).
FT METAL 85 85 IROn (HEME PROXIMAL LIGAND)
FT FT (BY SIMILARITY).
SQ NP_BIND 268 273 MADE (RIBOSE PART) (BY SIMILARITY).
SEQUENCE 396 AA; 43867 MW; 49961BDEL44ABD66 CRC64;

Query Match      2.8%; Score 7; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 EGGVSN 129
    |||||
Db 225 EGGVSN 231

RESULT 9
ID YDJN_HAEIN STANDARD; PRT; 440 AA.
AC P45079;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL SYMPORTER H1154.
GN H1154.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
CX NCBI_Taxid=727;
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Knelavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mc Kenney K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Georgagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1 SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF). STRONG, TO E.COLI YDJN.
CC CC
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between
CC CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL: U32795; AAC22809.1; -.
DR TIGR: H11154; -.
DR InterPro: IPR001991; -.
DR Pfam: PF00375; SDF; 1.
DR PROSITE: PS00713; NA_DICARBOXYL_SYM_1; FALSE NEG.
DR PROSITE: PS00714; NA_DICARBOXYL_SYM_2; FALSE NEG.
KW Hypothetical protein, Transmembrane, Transport.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
FT TRANSMEM 129 199 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
```

FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 366 386 POTENTIAL.
FT TRANSMEM 389 409 POTENTIAL.
SQ SEQUENCE 440 AA; 46491 MW; F9298893559PD06A CRC64;

Query Match 2.88; Score 7; DB 1; Length 440;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GOVSNTL 131
DB 150 GOVSNTL 156

RESULT 10
ETS2_MOUSE STANDARD; PRT; 468 AA.
AC P15037;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCN-2000 (Rel. 40, Last annotation update)
DE C-ETS-2 PROTEIN.
GN ETS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=89042086; PubMed=2847145;
RA Watson D.K., McWilliams M.J., Lapis P., Lautenberger J.A.,
RA Schweinkest C.W., Papes T.S.;
RA "Mammalian ets-1 and ets-2 genes encode highly conserved proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 85:7862-7866(1988).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J04103; AAA37581.1; -
CC PIR: C32066; TVMS2.
CC HSSP: P14921; 2STW.
CC TRANSFAC: T01397; -
CC MGD: MGI:95456; ETS2.
CC InterPro: IPR000418; -
CC Pfam: PF00178; Ets.1.
CC PRINTS: PR00454; ETSDOMAIN.
CC PROSITE: PS00345; ETS_DOMAIN_1; 1.
CC PROSITE: PS00346; ETS_DOMAIN_2; 1.
CC PROSITE: PS00061; ETS_DOMAIN_3; 1.
CC KM Proto-oncogene: DNA-binding; Nuclear protein.
FT DOMAIN 87 170 POINTED.
FT DNA_BIND 362 442 ETS_DOMAIN.
SQ SEQUENCE 468 AA; 52827 MW; 5260F3085B7E8831 CRC64;

Query Match 2.88; Score 7; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 190 FDFDGS 196
DB 29 FDFDGS 35

RESULT 11
ETS2_HUMAN STANDARD; PRT; 469 AA.
ID ETS2_HUMAN
AC P15036;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCN-2000 (Rel. 40, Last annotation update)
DE C-ETS-2 PROTEIN.
GN ETS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=89042086; PubMed=2847145;
RA Watson D.K., McWilliams M.J., Lapis P., Lautenberger J.A.,
RA Schweinkest C.W., Papes T.S.;
RA "Mammalian ets-1 and ets-2 genes encode highly conserved proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 85:7862-7866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Zimmermann W.W.K., Korenberg J., Rosenthal A., Schatteroy R.;
RA Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RL [3]
RX SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RX Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Seeda E.,
RX Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RX Menzel U., Delabar J., Kumpf K., Lehmann R., Paterson D.,
RX Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RX Rosenthal A., Kudoh T., Shibuya K., Kawasaki K., Asakawa S.,
RX Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RX Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RX Scharte M., Schoen O., Dessario A., Reichelt J., Kauer G., Bloeker H.,
RX Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RX Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RX Lehrich H., Reinhardt R., Vassio M.-L.;
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 399-469 FROM N.A.
RX MEDLINE=91067187; PubMed=2250910;
RX Watson D.K., Mavrothalassitis G.J., Jorcyk C.L., Smyth F.E.,
RA Papes T.S.;
RT "Molecular organization and differential polyadenylation sites of the
RT human ETS2 gene."
RL Oncogene 5:1521-1527(1990).
RN [5]
RP SEQUENCE OF 324-469 FROM N.A.
RX MEDLINE=86042652; PubMed=2997781;
RX Watson D.K., McWilliams-Smith M.J., Nunn M.F., Duesberg P.H.,
RA O'Brien S.D., Papes T.S.;
RT "The ets sequence from the transforming gene of avian
RT erythroblastosis virus, E26, has unique domains on human chromosomes
RT 11 and 21: both loci are transcriptionally active."
RL Proc. Natl. Acad. Sci. U.S.A. 82:7294-7298(1985).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J04102; AAA52412.1; -
CC DR EMBL: AF017257; AAB94057.1; -
CC EMBL: AP001732; BAA95514.1; -

Query Match 2.8%; Score 7; DB 1; Length 677;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ITGKAGA 80
DB 465 ITGKAGA 471

RESULT 14

MYSA_DICDI STANDARD; PRT; 994 AA.
ID MYSA_DICDI
AC P22467;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE MYOSIN IA HEAVY CHAIN (MYOSIN-LIKE PROTEIN ABMA).
GN ABMA OR DMIA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AS-2;
RX MEDLINE=92096486; PubMed=2519618;
RA Titus M.A., Warrick H.M., Spudich J.A.;
RT "Multiple actin-based motor genes in Dictyostelium.";
RL Cell Regul. 1:55-63(1989).

CC -1- FUNCTION: ACTIN-BASED MOTOR PROTEIN, POSSIBLY INVOLVED IN A WIDE
CC RANGE OF MOTILE PROCESSES, SUCH AS CELL MOVEMENT ACROSS A SURFACE,
CC AND EXTENSION AND RETRACTION OF PSEPODIA OR LAMELLIPODIA.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED, DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- SIMILARITY: BELONGS TO THE MYOSIN TYPE I FAMILY (SMALL MYOSINS).
CC -1- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: S73909; AAB20711.1; -;
CC PIR: J00151; J00151.
DR HSP: P08799; IMND.
DR DictyDb: DD01028; abma.
DR InterPro: IPR000048; -;
DR InterPro: IPR001609; -;
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF00063; myosin_head; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
KW Myosin; Actin-binding; ATP-binding; Multigene family.
FT DOMAIN 1 782 GLOBULAR HEAD-LIKE DOMAIN.
FT NP_BIND 783 994 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
FT NP_BIND 105 112 ATP (POTENTIAL).
FT DOMAIN 575 655 ACTIN-BINDING.
FT DOMAIN 830 931 MEMBRANE-BINDING (POTENTIAL).
SQ SEQUENCE 994 AA; 113286 MW; D30A5D2085B118C CRC64;

Query Match 2.8%; Score 7; DB 1; Length 994;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 EBYOYK 22
DB 694 EBYOYK 700

RESULT 15

SYIP-STAAU STANDARD; PRT; 1024 AA.
ID SYIP-STAAU
AC P41368;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE ISOLEUCYL-tRNA SYNTHETASE, MUPIROICIN RESISTANT (EC 6.1.1.5)
DE (ISOLEUCINE-tRNA LIGASE) (ILERS) (MUPIROICIN RESISTANCE PROTEIN).
GN MUPIR.
OS Staphylococcus aureus.
OC plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_Taxid=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J2870.
RX MEDLINE=94346838; PubMed=8067768;
RA Hodgson J.E., Curdock S.P., Dyke K.G.H., Morris R., Sylvester D.R.,
RA Gross M.S.;
RT "Molecular characterization of the gene encoding high-level mupirocin
RT resistance in Staphylococcus aureus J2870.";
RL Antimicrob. Agents Chemother. 38:1205-1208(1994).

CC [2]
CC SEQUENCE OF 504-519 AND 550-609 FROM N.A.
CC STRAIN=J2870;
CC MEDLINE=91244141; PubMed=1903747;
CC Dyke K.G.H., Curdock S.P., Golding M., Noble W.C.;
CC "Cloning of the gene conferring resistance to mupirocin in
CC Staphylococcus aureus.";
CC FEMS Microbiol. Lett. 61:195-198(1991).
CC -1- FUNCTION: CONFERS HIGH-LEVEL RESISTANCE TO THE ANTIBIOTIC
CC MUPIROICIN (PSEUDOMONIC ACID A), AN ILE-ANALOG THAT COMPETITIVELY
CC INHIBITS ACTIVATION BY ILE-tRNA SYNTHETASE THUS INHIBITING PROTEIN
CC BIOSYNTHESIS.
CC -1- CATALYTIC ACTIVITY: ATP + L-ISOLEUCINE + tRNA(ILE) = AMP +
CC PROPHOSPHATE + L-ISOLEUCYL-tRNA(ILE).
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X75439; CAAS3189.1; -;
CC EMBL: X59478; CAA42080.1; -;
CC EMBL: X59477; CAA42079.1; -;
DR InterPro: IPR001412; -;
DR InterPro: IPR002300; -;
DR InterPro: IPR002301; -;
DR Pfam: PF00133; tRNA-synt.1; 1.
DR PRINTS: PR00984; TRNASYNTHE.
KW Prosite: PS00178; AA-tRNA-LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Zinc; Antibiotic resistance; Plasmid.
FT SITE 42 52 "HIGH" REGION.
FT SITE 585 589 "KMSKS" REGION.
FT BINDING 588 588 ATP (BY SIMILARITY).
FT CONFLICT 519 519 P -> R (IN REF. 2).
FT CONFLICT 591 593 GNV -> ETE (IN REF. 2).
SQ SEQUENCE 1024 AA; 118875 MW; 56BCD232CA0C8430 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 1024;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RKQESF 236
| | | | |
Db 29 RKQESF 35

Search completed: August 23, 2001, 11:18:28
Job time: 195 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:13:38 ; Search time 19.66 Seconds

(without alignments)
432.114 Million full updates/sec

Title: US-09-654-652a-1

Perfect score: 1333

Sequence: 1 MVSADFGSGAEYLTLEVOY.....TRKGESFNGQVPRDEPAP 248

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1328	99.6	349 1 GUB_FIBSU	P17989 fibroblacter
2	266.5	20.0	259 1 GUB_BACBR	P37073 bacillus br
3	245.5	18.4	242 1 GUB_BACSU	P04957 bacillus su
4	242.5	18.2	802 1 XYND_RUMFL	O53317 rumiococcu
5	238.5	17.9	239 1 GUB_BACAM	P07980 bacillus am
6	237	17.8	334 1 GUB_CLOTM	P29716 clostridium
7	229.5	17.2	243 1 GUB_BACLI	P27051 bacillus li
8	226	17.0	237 1 GUB_PAEMA	P23904 paenibacilli
9	225	16.9	238 1 GUB_PAEPD	P43797 paenibacilli
10	197	14.8	269 1 EXOK_RHIME	P33693 rhizobium m
11	182.5	13.7	507 1 YG46_YEAST	P53301 saccharomyc
12	147	11.0	347 1 UTR2_YEAST	P24806 arabidopsis
13	147	11.0	347 1 UTR2_YEAST	P33623 saccharomyc
14	133.5	10.0	286 1 GUB_RHOMR	P45798 rhodothermu
15	128	9.6	283 1 BRU1_SOYBN	P33694 glycline max
16	117	8.6	397 1 CGKA_ALFCA	P43479 alteromonas
17	115	8.6	682 1 E13B_BACCI	P23903 bacillus ci
18	105	7.9	465 1 EXSH_RHIME	O33680 rhizobium m
19	102.5	7.7	879 1 GUNI_CLOTM	O02934 clostridium
20	102.5	7.7	1580 1 ACC8_HUMAN	O09428 homo sapien
21	95.5	7.2	1580 1 ACC8_HUMAN	O09429 rattus norv
22	95.5	7.2	1581 1 ACC8_CRICR	O09427 cricetus cr
23	94	7.1	465 1 EGIC_RHIME	O09342 rhatium m
24	91.5	6.9	571 1 XYNC_PSEFL	P23031 pseudomonas
25	90	6.8	490 1 ALGE_PSEAE	P18895 pseudomonas
26	88.5	6.6	736 1 VP4_ROTPE	P26193 porcine rot
27	87.5	6.6	400 1 GUN5_BACAG	O85465 bacillus ag
28	87	6.5	364 1 VM21_BORHE	P21875 borrelia he
29	86.5	6.5	1220 1 CSAC_BACUD	P56955 bacillus th
30	86.5	6.5	1385 1 CSAA_BACUD	O45760 bacillus th
31	84.5	6.3	409 1 GUN2_BACSD	P06565 bacillus sp
32	84.5	6.3	1289 1 CSAB_BACUD	O45753 bacillus th
33	83.5	6.3	422 1 FBW2_HUMAN	O9ukt8 homo sapien

34	82	6.2	422 1 FBW2_MOUSE	O60584 mus musculu
35	82	6.2	1103 1 CHS6_USTWA	O13395 ustiliago ma
36	81.5	6.1	614 1 BRUB_ECOCI	P06129 escherichia
37	81	6.1	411 1 DEOB_LACIC	O32808 lactococcus
38	81	6.1	551 1 AMT4_PSESA	P22963 pseudomonas
39	80.5	6.0	708 1 PAL1_DAUCA	O23865 daucus caro
40	79.5	6.0	571 1 PAL4_PORPI	O40910 populus kit
41	79.5	6.0	2628 1 HAGA_PORGI	O51845 porphyromon
42	79	5.9	1042 1 EFG_PNECA	P29551 pneumocysti
43	78.5	5.9	301 1 OMPG_ECOCI	P76045 escherichia
44	78.5	5.9	449 1 TBB2_EMENT	P10874 emericella
45	78.5	5.9	464 1 ACHO_RAT	P12391 rattus norv

ALIGNMENTS

RESULT	ID	STANDARD	PRT	349 AA.
1	GUB_FIBSU			
AC	P17989:			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)			
DE	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).			
OS	Fibroblacter succinogenes (Bacteroides succinogenes).			
OC	Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group; Fibrobacter.			
OC	NCBI_TaxID=833;			
OX	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.			
RC	STRAIN-ISOLATE S85;			
RX	MEDLINE-90299807; PubMed-2193918;			
RA	Teather R.M., Erle J.D.;			
RT	"DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";			
RL	J. Bacteriol. 172:3837-3841(1990).			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: M33676; AAA24896.1;			
DR	PIR: A44507; A44507.			
DR	HSSP: P23904; IAJK.			
DR	InterPro: IPR000757;			
DR	Pfam: PF00722; Glyco_hydro_16; 1.			
DR	PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.			
DR	Hydrolase; Glycosidase; Signal; Repeat.			
FT	SIGNAL	1	27	
FT	CHAIN	28	349	
FT	ACT_SITE	79	79	BETA-GLUCANASE.
FT	ACT_SITE	83	83	NUCLEOPHILE (BY SIMILARITY).
FT	DOMAIN	271	307	PROTON DONOR (BY SIMILARITY).
FT	REPEAT	271	277	5 X 7 AA TANDEN REPEATS OF P-X-S-S-S-X.
FT	REPEAT	278	284	1.
FT	REPEAT	285	291	2.
FT	REPEAT	292	298	3.
FT	REPEAT	301	307	4.
SO	SEQUENCE	349 AA;	37737 MM;	16DC4F5BDEFC578A CRC64;

Query Match 99.6%; Score 1328; DB 1; Length 349;

Best Local Similarity 100.0%; Pred. No. 5e-104;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VSADFGAELETTLEEVYQGFKEARMKMAASGVSSMFLYONSEIADGRPWVEVDIEV 61
 |||||
 Db 25 VSADFGAELETTLEEVYQGFKEARMKMAASGVSSMFLYONSEIADGRPWVEVDIEV 84
 |||||
 QY 62 LKNGPSFQSNITITGKAGAKTSEKHNAVSPADAQFHTYGLEWTPYVRYVTVDGQEVK 121
 |||||
 Db 85 LKNGPSFQSNITITGKAGAKTSEKHNAVSPADAQFHTYGLEWTPYVRYVTVDGQEVK 144
 |||||
 QY 122 TEGGVSNLTGTOGLRNRNMSSESAAMVQGFDESKLPFOFINMKYKKTTPGQEGGSD 181
 |||||
 Db 145 TEGGVSNLTGTOGLRNRNMSSESAAMVQGFDESKLPFOFINMKYKKTTPGQEGGSD 204
 |||||
 QY 182 FTLDWTDNFTPDGSRMGKGMDFDGNRVLDTDKNISRDGMLLALTRKGOSEFNGOVP 241
 |||||
 Db 205 FTLDWTDNFTPDGSRMGKGMDFDGNRVLDTDKNISRDGMLLALTRKGOSEFNGOVP 264
 |||||
 QY 242 RDDEPAP 248
 |||||
 Db 265 RDDEPAP 271
 |||||

RESULT 2

GDB_BACBR STANDARD; PRT; 259 AA.

AC P37073;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLBB.
 OS Bacillus brevis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Brevibacillus.
 OX NCBI_TaxID=1393;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93159752; PubMed=7763386;
 RA Louw M.E., Reid S.J., Watson T.G.;
 RT "Characterization, cloning and sequencing of a thermostable
 endo-(1,3-1,4) beta-glucanase-encoding gene from an alkalophilic
 RT Bacillus brevis.";
 RT Appl. Microbiol. Biotechnol. 38:507-513(1993).
 CC -1- FUNCTION: HYDROLYSES B-GLUCANS CONTAINING MIXED BETA-1,3 AND
 BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; ITS OPTIMAL
 TEMPERATURE IS NEAR 65-70 DEGREES CELSIUS.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC or send an email to license@isb-sib.ch).
 CC EMBL: M84339; AAA22265.1; -
 DR HSSP: P23904; IAJK.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS: PR00737; GLHYDRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolyase; Glycosidase; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 259 POTENTIAL.
 FT ACT_SITE 142 259 BETA-GLUCANASE.
 FT ACT_SITE 146 146 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 146 146 PROTON DONOR (BY SIMILARITY).

SQ SEQUENCE 259 AA; 29960 MW; A63C09F281FF5D13 CRC64;

Query Match 20.0%; Score 266.5; DB 1; Length 259;
 Best Local Similarity 36.08; Pred. No. 26-15;
 Matches 63; Conservative 24; Mismatches 75; Indels 13; Gaps 6;

QY 3 SAKDFSGAELETTLEEVYQGFKEARMKMAASGVSSMFLYONSEIADGRPWVEVDIEV 62
 |||||
 Db 90 SAKRYKAGELRTNDFHYHGLFEVSMKPAKVEGYSSFFTY-TGEMWDGDPMPDEIDIEFL 148
 |||||
 QY 63 GKNGPSFQSNITITGKAGAKTSEKHNAVSPADAQFHTYGLEWTPYVRYVTVDGQEVK 122
 |||||
 Db 149 GKDTTRIQFNFTYVTVGVC---NEFYDLDGFDASESEFNTYAFEMREDSITVYVNGEAVHTA 205
 |||||
 QY 123 EGGGVSNLTGT-OGLRNRNMSSESA-AMVQGFDESKLPFOFINMKYKKTTPGQ 175
 |||||
 Db 206 ---TENIPQTPOKIMNMIMPVGVDGTVGFDGNDTPVSYXYDW---RYTPIQ 253
 |||||

RESULT 3

GDB_BACSU STANDARD; PRT; 242 AA.

AC P04957;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLS OR BGL OR LICS OR N15B.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C120;
 RX MEDLINE=84272222; PubMed=6087283;
 RA Murphy N., McConnell D.J., Cantwell B.A.;
 RT "The DNA sequence of the gene and genetic control sites for the
 RT excreted B. subtilis enzyme beta-glucanase.";
 RT Nucleic Acids Res. 12:5355-5367(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HD-25;
 RA Tezuka H., Yunki T., Yabuuchi S.;
 RT "Construction of a beta-glucanase hyperproducing Bacillus subtilis
 RT using the cloned beta-glucanase gene and a multi-copy plasmid.";
 RT Agric. Biol. Chem. 53:2335-2339(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSJAL;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95219081; PubMed=7704256;
 RA Wolf M., Geczi A., Simon O., Boriss R.;
 RT "Genes encoding xylan and beta-glucan hydrolysing enzymes in Bacillus
 RT subtilis: Characterization, mapping and construction of strains
 RT deficient in lichenase, cellulase and xylanase.";
 RT Microbiology 141:281-290(1995).
 RN [5]
 RP SEQUENCE OF 1-6 FROM N.A.
 RC STRAIN=BR151;
 RX MEDLINE=96178961; PubMed=8606172;
 RA Schmetz K., Stuelke J., Gertz S., Krueger S., Krieg M., Hecker M.,
 RA Rak B.;
 RT "LicB, a Bacillus subtilis transcriptional antiterminator protein of
 RT the BglG family.";
 RN J. Bacteriol. 178:1971-1979(1996).
 RN [6]

```

RA SEQUENCE OF 29-63.
RT Yuuki T., Tezuka H., Yabuchi S.:
RT "Purification and some properties of two enzymes from a beta-glucanase
RT hyperproducing strain, Bacillus subtilis HL-25."
RL Agric. Biol. Chem. 53:2341-2346(1989).
CC -1- CATALYTIC ACTIVITY: HYDROLAYS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC -1- SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- MISCELLANEOUS: THE SEQUENCE OF STRAIN 168 IS SHOWN.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00754; CA25328.1; -
DR EMBL; D00518; BAA00405.1; -
DR EMBL; D83026; BAA11697.1; -
DR EMBL; Z46862; CA86822.1; -
DR EMBL; Z28340; CA82195.1; -
DR EMBL; Z99124; CAB15943.1; -
DR PIR; A22914; LXBS.
DR PIR; J00110; J00110.
DR HSSP; P27051; 1GBG.
DR Subtilist; BGL0476; bg1S.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KM Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 28
FT CHAIN 29 242
FT MOD_RES 29 29
FT ACT_SITE 133 133
FT ACT_SITE 137 137
FT DISULFID 60 89
FT VARIANT 24 24
FT VARIANT 83 83
FT VARIANT 204 204
FT VARIANT 204 204
SQ SEQUENCE 242 AA; 27268 MW; 45958DEA0F22B29 CRC64;

Query Match 18.4%; Score 245.5; DB 1; Length 242;
Best Local Similarity 35.8%; Pred. No. 1,1e-13;
Matches 59; Conservative 24; Mismatches 71; Indels 11; Gaps 6;

OY 7 FSGAEIYLEEVQYKFPARKMAASGVSSMFLYONGSEFADGRPVEVDIEVLGNP 66
Db 87 FDCCGNRSVQYRYGGLYEVRRMKRPANNTGIVSFFTYPT--DETPEDEIDIEFLGDT 143
OY 67 GSFOSNITITGGAQAKTSEKHAAVSPADQAFHTYGLCEWTPYVYVWVDGQEVKTEGQ 126
Db 144 TKVQENYTTNGA---NHEKIYDLGFDAAANAHHTYAFQMOPNSIMWYDGO LKHTATNQ 199
OY 127 VSNLTGTGLRFLNMSSEA-AWVGQFDESKLPILQFTINWVKVY 170
Db 200 IPTPGK--IMMINLWNGTGVDEMLGSYNGVN-PLVAHYHDWRYTK 241

RESULT 4
XYND_RUMFL STANDARD; PRT; 802 AA.
AC 053117;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE XYLANSE/BETA-GLUCANASE PRECURSOR [INCLUDES: ENDO-1,4-BETA-XYLANASE

```

DE (EC 3.2.1.8) (XYLANASE); (ENDO-BETA-1,3-1,4 GLUCANASE) (EC 3.2.1.73)
 EC [1,3-1,4-BETA-D-GLOCAN 4-GLUCANOHYDROLASE] (LICHEANASE)].
 GN XYND.
 GN Ruminococcus flavefaciens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Ruminococcus.
 OC NCBI_TaxID=1265;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17;
 RC MEDLINE=93259938; PubMed=8491715;
 RA Flint H.J., Martin J., McPherson C.A., Daniel A.S., Zhang J.-X.;
 RT "A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-
 RT glucanase domains, encoded by the xynd gene of Ruminococcus
 RT flavefaciens.";
 RT J. Bacteriol. 175:2943-2951(1993).
 RL -I- FUNCTION: CONTAINS TWO CATALYTIC DOMAINS WITH XYLANASE AND ENDO-
 CC BETA-1,3-1,4 GLUCANASE ACTIVITIES.
 CC -I- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
 CC LINKAGES IN XYLANS.
 CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTRAINING 1,3- AND 1,4-BONDS.
 CC -I- PATHWAY: XYLAN DEGRADATION.
 CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC G (FAMILY 11 OF GLYCOSYL HYDROLASES).
 CC -I- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 16 OF
 CC GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/announce/
 CC or_send_an_email_to_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch)).
 CC -----
 DR EMBL; S61204; AAB26620.1; -;
 DR HSSP; P23904; IABK.
 DR Interpro; IPR000757; -;
 DR Interpro; IPR001137; -;
 DR Pfam; PF02018; CBD_6; 1.
 DR Pfam; PF00457; Glyco_hydro_11; 1.
 DR Pfam; PF00722; Glyco_hydro_16; 1.
 DR PRINTS; PR00737; GLHYDRASE16.
 DR PRINTS; PR00911; GLHYDRASE11.
 DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
 DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 DR KW Xylan degradation; Hydrolase; Glycosidase; Signal;
 KM Multifunctional enzyme.
 FT SIGNAL 1 31
 FT CHAIN 32 802
 FT DOMAIN 32 244
 FT DOMAIN 245 523
 FT DOMAIN 524 555
 FT DOMAIN 556 802
 FT ACET_SITE 124 124
 FT ACET_SITE 226 226
 FT ACT_SITE 684 684
 FT DOMAIN 524 529
 FT DOMAIN 532 543
 FT DOMAIN 546 553
 FT POLY-THR.
 FT POLY-THR.
 SQ SEQUENCE 802 AA; 89091 MM; 2880A689647284AF CRC64;

DR	EMBL; X63355; CAA44959.1; -
DR	EMBL; X58392; CAA11281.1; -
DR	PIR; S23498; S23498
DR	PIR; J05611; J50611
DR	PIR; S18726; S18726
DR	HSSP; P23904; 1A9K
DR	InterPro; IPR000757; -
DR	InterPro; IPR002048; -
DR	InterPro; IPR002105; -
DR	Pfam; PF00404; Dockerin_1; 2.

DB 141 GKDDTKVOFNYYTNGVGV---NHEKIVNLGFDPAANSYHTYAFDQMPNSIKWVDGO-LKHT 196
 OY 123 EGGGVSNLTGQGRFNLMSSESA-AMVGQFDESKLPFQFINWVKYK 170
 DB 197 ATTQIPOTPGK--IMMNLNGAGVDEWLGSYN-GVTPLSRSLHWRYTK 242

RESULT 8
 GUB_PAEPA STANDARD: PRT: 237 AA.

AC P23904;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 OS Paenibacillus macerans (Bacillus macerans).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Paenibacillus.
 OX NCBI_TaxID=44252;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91109712; PubMed=2274030;
 RA Borris R., Bueltner K., Maentzsaellae P.;
 RT "Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
 RT homologies to other beta-glucanases.";
 RL Mol. Genet. 222:278-283(1990).
 RN [12]
 RP ACTIVE SITE
 RX MEDLINE=93094208; PubMed=1360982;
 RA Hoel J.P., Condon R., Traeger J.C., Stone B.A.;
 RT "Identification of glutamic acid 105 at the active site of Bacillus
 RT amyloliquefaciens 1,3-1,4-beta-D-glucan 4-glucanohydrolase using
 RT epoxide-based inhibitors.";
 RL J. Biol. Chem. 267:25059-25066(1992).
 RN [13]
 RP MUTAGENESIS OF GLU-128.
 RA Olsen O.;
 RL Thesis (1990), University of Aarhus, Denmark.
 RN [14]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=93281743; PubMed=8099449;
 RA Kettel T., Simon O., Borris R., Heinemann U.;
 RT "Molecular and active-site structure of a Bacillus 1,3-1,4-beta-
 RT glucanase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5287-5291(1993).
 RN [15]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX Hahn M., Kettel T., Heinemann U.;
 RT MEDLINE=96028129; PubMed=7588726;
 RT "Crystal and molecular structure at 0.16-nm resolution of the hybrid
 RT Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-W).";
 RL Eur. J. Biochem. 232:849-858(1995).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X55959; CAA39426.1;
 DR PIR: S11927; S11927.
 DR PDB: ZATH; 31-MAR-95.
 DR PDB: BYH; 31-OCT-93.
 DR PDB: ICPM; 22-JUN-94.

DR PDB: ICPM; 22-JUN-94.
 DR PDB: IMAC; 27-FEB-95.
 DR PDB: LAJK; 06-MAY-98.
 DR PDB: LAJO; 06-MAY-98.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS: PR00737; GLHYDRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal; 3D-structure.
 FT SIGNAL 1 23
 FT CHAIN 24 237
 FT ACT_SITE 128 128
 FT ACT_SITE 132 132
 FT DISULFID 55 84
 FT MUTAGEN 128 128
 FT STRAND 26 26
 FT STRAND 29 31
 FT STRAND 41 43
 FT TURN 51 52
 FT STRAND 53 53
 FT STRAND 55 57
 FT STRAND 56 61
 FT HELIX 59 61
 FT STRAND 62 64
 FT TURN 66 67
 FT STRAND 70 78
 FT TURN 79 80
 FT STRAND 81 89
 FT STRAND 93 93
 FT STRAND 96 103
 FT TURN 108 109
 FT STRAND 110 118
 FT HELIX 120 122
 FT TURN 123 123
 FT STRAND 127 134
 FT TURN 135 136
 FT STRAND 137 147
 FT TURN 148 149
 FT STRAND 150 150
 FT STRAND 155 158
 FT TURN 163 164
 FT STRAND 168 174
 FT STRAND 179 183
 FT TURN 184 185
 FT STRAND 186 191
 FT TURN 200 209
 FT TURN 213 216
 FT STRAND 225 236
 SQ SEQUENCE 237 AA; 26589 MW; 436EABCDFFC87781 CRC64;

Query Match 17.0%; Score 226; DB 1; Length 237;
 Best Local Similarity 35.1%; Pred. No. 4, 4e-12;
 Matches 59; Conservative 23; Mismatches 70; Indels 16; Gaps 8;

OY 7 FSGAEVLTLEEVYQKFEKFAKMAAGTVSSMFLYQNGSEIADGPRVVEVIEVLGKRP 66
 DB 82 FDCAEFRSINIIYGYGLXYSMPKAKNTGIVSSFTYTG---AHGQNDIEDIEFLGKRT 138
 OY 67 GSFQSMIITGKAGAOQKTSKHNHVAQPAQDAFHVTYGLTEMTPNVRYVTVDGQEVKRTGGO 126
 DB 139 TKVQFVNYVINGVGH---EKVSLSGFDASKGFHYTAFDQPGYIKWYVDG-VLKHT---A 191
 OY 127 VSNLTGTGQ-LRFNLMSSESA-AMVGQFDESKLPFQFINWVKYK 172
 DB 192 TANIPSTPGKIMMNLNGAGVDEWLGSYNGAN-PLVAEYDWV---KYT 235

RESULT 9
 GUB_PAEPO STANDARD: PRT: 238 AA.
 ID GUB_PAEPO
 AC P45797;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN GLUB.
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 842.
RC MEDLINE=92041687; PubMed=1938968;
RA Gosalbes M.J., Perez-Gonzalez J.A., Gonzalez R., Navarro A.;
RT "Two beta-glucanase genes are clustered in Bacillus polymyxa:
RT molecular cloning, expression, and sequence analysis of genes
RT encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
RL J. Bacteriol. 173:7705-7710(1991).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57094; CAA40379.1; -
DR HSSP: P23904; ICPM.
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLAHDRLASE16.
DR PROSITE: PS01034; GLYCOSYL-HYDROL_F16; 1.
DR KMW: KMW000000; Glycosylase; Signal.
FT SIGNAL 1 26
FT CHAIN 27 238
FT ACT_SITE 129 129 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 133 133 PROTON DONOR (BY SIMILARITY).
FT DISULFID 56 85 BY SIMILARITY.
SQ SEQUENCE 238 AA; 26919 MW; C0CF7B4E5D40E8C CRC64;
Query Match 16.9%; Score 225; DB 1; Length 238;
Best Local Similarity 34.5%; Pred. No. 5.4e-12;
Matches 58; Conservative 23; Mismatches 71; Indels 16; Gaps 8;
OY 7 FSGAEIYTLIEVOYGFEPARKMAASGVSSMFLYONGSEIADGRPRVVEVDIEVLGKNP 66
DB 83 FDCEIYSTNNYGYGLYEVSKPKAKNTGIYSSFTYTGPSH---GTOWDEIDIEFLGKDT 139
OY 67 GSPFSNITTKAGAKOAKTSEKHHAVSPADQAFHTYGLTEPNTYVWTVGQGEVKTESGQ 126
DB 140 TKVOFNYYTNGVGH---EKTIINIGPDASTSFHTYAFPMWGGYIKWYDG-VLAKHT---A 192
OY 127 VSNLTGQG-LRFNLMSSESA-AWVGQFDESKLPLPQFINNKKYKKT 172
DB 193 TTNIPTSPGKIMMIMNGTGVSWLGSYGAN-PLXAEYDMV---KYT 236
RESULT 10
EXOK_RHIME STANDARD; PRT; 269 AA.
AC P33693;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENDO-1,3-1,4-BETA-GLUCANASE EXOK PRECURSOR (EC 3.2.1.-) (SUCCINOGLYCAN
DE BIOSYNTHESIS PROTEIN EXOK).
GN EXOK.

OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RC MEDLINE=94042869; PubMed=8226645;
RA Gluckmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT succinoglycan by Rhizobium meliloti.";
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=RCR2011 / SU47;
RC MEDLINE=93241147; PubMed=8479421.
RA Becker A., Kleckmann A., Arnold W., Puhler A.;
RT "Analysis of the Rhizobium meliloti exoH/exoK/exoL fragment: ExoK
RT shows homology to excreted endo-beta-1,3-1,4-glucanases and ExoH
RT resembles membrane proteins.";
RL Mol. Gen. Genet. 238:145-154(1993).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98226741; PubMed=9560202;
RA York G.M., Walker G.C.;
RT "The Rhizobium meliloti ExoK and ExoH glucanases specifically
RT depolymerize nascent succinoglycan chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:4912-4917(1998).
CC -1- FUNCTION: CLEAVES HIGH MOLECULAR WEIGHT SUCCINOGLYCAN TO YIELD LMW
CC SUCCINOGLYCAN. DYNAMICALLY REGULATE THE MOLECULAR WEIGHT
CC DISTRIBUTION OF SUCCINOGLYCAN BY CLEAVING NASCENT SUCCINOGLYCAN
CC ONLY DURING A LIMITED PERIOD AFTER ITS SYNTHESIS, PERHAPS BEFORE
CC IT UNDERGOES A TIME-DEPENDENT CHANGE IN ITS CONFORMATION OR
CC AGGREGATION STATE.
CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L20758; AAA16048.1; -
DR EMBL: Z17219; CAA78927.1; -
DR HSSP: P23904; IAKJ.
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PROSITE: PS01034; GLYCOSYL-HYDROL_F16; 1.
DR KMW: KMW000000; Glycosidase; Hydrolase; Signal; Plasmid.
FT SIGNAL 1 27
FT CHAIN 28 269
FT CONFLICT 93 102 ACCEIOTRKR -> LRNPDAOG (IN REF. 2).
SQ SEQUENCE 269 AA; 30083 MW; 073CC7ED65E2611 CRC64;
Query Match 14.8%; Score 197; DB 1; Length 269;
Best Local Similarity 32.5%; Pred. No. 1.4e-09;
Matches 50; Conservative 31; Mismatches 61; Indels 12; Gaps 6;
OY 2 VSAKDFGAEIYTLIEVOYGFEPARKMAASGVSSMFLYONGSEIADGRPRVVEVDIEV 61
DB 87 VKERNFACGEIOTRKRREYGYTEYERIKAADSGNSAFYTYGP---ADKRPHEIDIEV 143
OY 62 LGKNRFSQSN-IITGKAGAKOAKTSEKHHAVSPADQAFHTYGLTEPNTYVWTVGQGEV 120
DB 144 LGKNTAKQIINOYVSAAKGNELAD---VPGANGQNDYAFWEKRIYRYNGELVH 199
OY 121 K-TEGGVYSNLTGTQGLRFLNLMSSESA-AWVGQF 152

DB 200 EVTDPAKIP--VNAOKIFFSLMGTDLTDMGTF 231

RESULT 11
ID YG46_YEAST STANDARD; PRT; 507 AA.

AC P53301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOHETICAL 52.8 KDA PROTEIN IN BUB1-HIPI INTERGENIC REGION.
GN YGRI89C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nombela C.;
RT DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII.";
RL Yeast 13:357-363(1997).
CC -1- SIMILARITY: SOME, TO YEAST UTR2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 272974; CAA97215.1; -
CC DR EMBL: X99074; CAA67525.1; -
CC DR HSSP: P23904; IAO.
CC DR SGD: S0003421; CRH1.
CC DR InterPro: IPR000757; -
CC DR Pfam: PF00722; Glyco_hydro_16; 1.
CC DR Hypothetical protein.
CC FT DOMAIN 63 66 POLY-SER.
CC FT DOMAIN 301 310 POLY-SER.
CC FT DOMAIN 345 357 POLY-SER.
CC FT DOMAIN 387 391 POLY-SER.
CC FT DOMAIN 467 470 POLY-SER.
CC SQ SEQUENCE 507 AA: 52757 MW: 7D7B61F57AE942C CRC64;

Query Match 13.7%; Score 182.5; DB 1; Length 507;
Best Local Similarity 26.7%; Pred. No. 4.7e-08;
Matches 58; Conservative 37; Mismatches 91; Indels 31; Gaps 8;

QY 1 MWSAKDFSGAELTYLLEVOYGFKAARKMAAAGTSSMFLYONGSEIADGRPWVEVDIE 60
DB 87 MTLAKRYDNPSPKSNFYIMGKLEVLKANGIIVSFFLQSDDD-----EIDIE 138
QY 61 VLAKNGSFQSNITITGKAGAKTSEKHAAVSPADQAFHTYGLEWTPNRYRWTVQOEVR 120
DB 139 WVGGDWTFQFSNFFSGDFTTYDRGEFHGVDPDPTDK-FHNHTLDMAMDKTWTWLDGESVR 197
QY 121 KTEGGOVSNLTGQGL-----LRFNLMS-----ESAA-----WV-GQFDESKLPLFOFINW 165
DB 198 -----VLSTSESGYPOSPMYLMGIMAGDPDANAGTIEMAGETNNVNDAPFTWIEK 251
QY 166 VKVYKTYTPGGESESDFTLMDTNTDFDGSRWKGD 202
DB 252 VIATDSTGKKYTYGDSGSW-ESIEADGSGTYGRD 287

RESULT 12

MERS_ARATH
ID MERS_ARATH STANDARD; PRT; 247 AA.

AC P24806;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE MERS-5 PROTEIN.
GN MERS-5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93005704; PubMed=1840916;
RA Medford J.I., Elmer J.S., Klee H.J.;
RT "Molecular cloning and characterization of genes expressed in shoot
RT apical meristems.";
RL Plant Cell 3:359-370(1991).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN SHOOT APICAL MERISTEMS, ALSO
CC FOUND IN SEEDLINGS AND MERISTEMS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M63166; AAA32828.1; -
CC DR PIR: J01022; J01022.
CC DR HSSP: P23904; IATK.
CC DR InterPro: IPR000757; -
CC DR Pfam: PF00722; Glyco_hydro_16; 1.
CC SQ SEQUENCE 247 AA: 28295 MW: 831E8441564B45E8 CRC64;

Query Match 11.0%; Score 147; DB 1; Length 247;
Best Local Similarity 28.0%; Pred. No. 1.9e-05;
Matches 44; Conservative 29; Mismatches 60; Indels 24; Gaps 7;

QY 8 SGAEELTYLLEVOYGFKAARKMAA--ASGTSSMFLYONGSEIADGRPWVEVDIEVLGK- 64
DB 54 SSGSFQSKTYLFEKIDMQIKLVPAGNSAGVTTFYIAKSEGS-----TWDEIDFEFLGMN 107
QY 65 --NPGSFQSNITITGKAGAKTSEKHAAVSPADQAFHTYGLEWTPNRYRWTVQDGO--EVR 120
DB 108 SGDPYTLHTWTVYQKGG-DKEQGFHLMFDEPTAN--PFTYSILMNPORILTVDTHPLES 164
QY 121 KTEGGOVSNLTGQGLR-----NLMSSESAANV 149
DB 165 KTWSLSVSCFQRTKPRMVRQFMEROSDWATKRSM 201

RESULT 13

ID UTR2_YEAST STANDARD; PRT; 347 AA.

AC P32623;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE UTR2 PROTEIN (UNKNOWN TRANSCRIPT 2 PROTEIN).
GN UTR2 OR YEL040W OR STGP-ORF18.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;

RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunticke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.,
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B-6441;
 RA MEDLINE=94016558; PubMed=8411151;
 RA Melnick L., Sherman F.;
 RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
 RT of Saccharomyces cerevisiae share a common ancestry."
 RL J. Mol. Biol. 233:372-388(1993).
 CC -1- SIMILARITY: SOME, TO YEAST YGR189C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL; U18779; AAB65002.1; ALT_INIT.
 DR EMBL; L22173; AAA34941.1; -
 DR EMBL; S65964; AAD13975.1; -
 DR EMBL; S66130; AAB28444.1; -
 DR PIR; S30839; S30839.
 DR HSSP; P23804; IAKO.
 DR SGD; S0000766; UTR2.
 DR InterPro; IPR000757; -
 DR Pfam; PF00722; Glyco_hydro_16; 1.
 DR DOMAIN 234 322 SER-RICH.
 FT DOMAIN 269 283 POLY-SER.
 FT CONFLICT 10 10 L->V (IN REF. 3).
 FT CONFLICT 171 171 A->R (IN REF. 3).
 FT CONFLICT 234 234 S->C (IN REF. 3).
 FT SEQUENCE 347 AA; 36692 MW; 1E8AFB862C4BB328 CRC64;
 SQ
 Query Match 11.0%; Score 147; DB 1; Length 347;
 Best Local Similarity 25.1%; Pred. No. 2.8e-05;
 Matches 50; Conservative 30; Mismatches 71; Indels 48; Gaps 9;
 QY 5 KDFSGAEIYLTLEEVQYGFKAARKMAASGVYSSMFLYQNGSEIADGRPWVEVDIEVLGK 64
 DB 3 KNSGGTIVSSYFRAVWYGVSAIKITSLAGVYTGFIISGACD-----ELDIEFVGA 54
 QY 65 NPGSFOSNI-----ITGKGAQOKTSEKHAAVSPADAQAFHTYGLGEMTPNVYRTVDOQ 117
 DB 55 DLEFQOTIFMYESVYANTNSANISITD-----TFENHTYELDHEDYVTSIDGV 105
 QY 118 EVRKTEGGOVSNLTGQGLRF-----NLW-----SSEA-----AWG-----QFDESKLP 158
 DB 106 VGRILYKRETYNAT-TQKXQYQGPSPKVDISIMPGNSTNAGTIAWSGGGEIMNDASDIS 164
 QY 159 -----LFOFINWVKYKYP 173
 DB 165 NPGYYVAIVNEVNTICYP 183
 RESULT 14
 GUB_RHOMR STANDARD; PRT; 286 AA.
 ID GUB_RHOMR
 AC P45798;
 DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGIA.
 OS Rhodothermus marinus.
 OC Bacteria; CFB group; Rhodothermus group; Rhodothermus.
 OX NCBI_TaxID=29549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-21 / ITI-378;
 RX MEDLINE=95010084; PubMed=7925416;
 RA Spilliaert R., Hreggvidsson G.O., Kristjansson J.K.,
 RA Eggertsson G., Palsdottir A.;
 RT "Cloning and sequencing of a Rhodothermus marinus gene, bgia, coding
 RT for a thermostable beta-glucanase and its expression in Escherichia
 RT coli."
 RL Eur. J. Biochem. 224:923-930(1994).
 CC -1- FUNCTION: SHOWS ACTIVITY ON LICHENAN, BETA-GLUCAN AND LAMINARIN
 CC BUT NOT ON CMC CELLULOSE OR XLAN. THE ENZYME HAS A TEMPERATURE
 CC OPTIMUM OF 85 DEGREES CELSIUS AND A PH OPTIMUM OF 7.0.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL; U04836; AAA60459.1; -
 DR HSSP; P23904; IAKR.
 DR InterPro; IPR000757; -
 DR Pfam; PF00722; Glyco_hydro_16; 1.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 DR KW Hydrolyase; Glycosidase; Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 286 BETA-GLUCANASE.
 FT ACT_SITE 158 158 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
 FT SEQUENCE 286 AA; 33145 MW; 7215C33624135191 CRC64;
 SQ
 Query Match 10.0%; Score 133.5; DB 1; Length 286;
 Best Local Similarity 23.7%; Pred. No. 0.0003;
 Matches 44; Conservative 34; Mismatches 85; Indels 23; Gaps 8;
 QY 5 KDFSGAEIYLTLEEVQYGFKAARKMAASGVYSSMFLYQNGSEIADGRPWV---EVDI 59
 DB 103 KEYTSARILTRGKASWYIGREFIRARLPBSGRGTWPAIMMLPRQTYGSAY-WPDNGEIDI 161
 QY 60 -EVLGKNP-----GSFOSNIITGKGAQOKTSEKHAAVSPADAQAFHTYGLGEMTPNVYRTV 114
 DB 162 MEHVGFNPDVVAGTYHTAYNHLSTQRCGLR---VPTARDPFIYVAIEWPEEIRMY 218
 QY 115 DQGEVARKTEGGOVSNLTG-----TQGLRFNLWSSSEAAWVG---FDESKLPFLQFINW 165
 DB 219 DDLTYREFPNERLTLPEDAWRHMPDPFPHILMNTIAVAGAMGQGVDPPEARPAULVDY 278
 QY 166 VVYVYK 171
 DB 279 VAVYRW 284
 RESULT 15
 BRU1_SOYBN STANDARD; PRT; 283 AA.
 ID BRU1_SOYBN
 AC P35694;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE BRASSINOSTEROID-REGULATED PROTEIN BR1.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Eudicotyledons; Core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Glycine.
 OX NCBI_TaxId=3847;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=epicotyl;
 RC MEDLINE=94159788; PubMed=8115544;
 RX Zurek D M., Clouse S.D.,
 RA "Molecular cloning and characterization of a
 RT brassinosteroid-regulated gene from elongating soybean (Glycine max
 RL L.) epicotyls."
 RL Plant Physiol. 104:161-170(1994).
 CC -1- FUNCTION: POSSIBLE ROLE IN BRASSINOSTEROID-STIMULATED ELONGATION.
 CC -1- SIMILARITY: TO A.THALIANA MER1-5.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L22162; AAA81350.1; -
 DR HSSP: P23904; IAN7.
 DR Interpro: IPR000577; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 SO SEQUENCE 283 AA; 32254 MW; C2488810EC7835737 CRC64;

Query Match 9.6%; Score 128; DB 1; Length 283;

Best Local Similarity	23.7%;	Pred.No.	0.00085;
Matches	46;	Conservative	41;
		Mismatches	73;
		Indels	34;
		Gaps	9;

[illegible]

```
Search completed: August 23, 2001, 11:16:30
Job time: 172 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 13:37:21 ; Search time 2905.14 Seconds
(without alignments)
4280.713 Million cell updates/sec

Title: US-09-654-652a-5
Perfect score: 804
Sequence: 1 atggttagcgcaagaattt.....accacgacacacccactga 804

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_da1:*
17: em_da2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_v13:*
61: gb_v14:*
62: gb_v15:*
63: gb_v16:*
64: gb_v17:*
65: gb_v18:*
66: gb_v19:*
67: gb_v20:*
68: gb_v21:*
69: gb_v22:*
70: gb_v23:*
71: gb_v24:*
72: gb_v25:*
73: gb_v26:*
74: gb_v27:*
75: gb_v28:*
76: gb_v29:*
77: gb_v30:*
78: gb_v31:*
79: gb_v32:*
80: gb_v33:*
81: gb_v34:*
82: gb_v35:*
83: gb_v36:*
84: gb_v37:*
85: gb_v38:*
86: gb_v39:*
87: gb_v40:*
88: gb_v41:*
89: gb_v42:*
90: gb_v43:*
91: gb_v44:*
92: gb_v45:*
93: gb_v46:*
94: gb_v47:*
95: gb_v48:*
96: gb_v49:*
97: gb_v50:*
98: em_da3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	742.8	92.4	1426	3	FTBGLUC
2	64.8	8.1	4411	9	AR069366
3	64.8	8.1	5443	9	AR069362
4	64.8	8.1	5616	9	AR069364
5	61.4	7.6	2735	9	AR054309
6	55	6.8	6363	9	AR069367
7	50	6.2	1416	2	BSPBGAIG
8	49.6	6.2	1234	56	SYHA78M

[illegible]

Db 937 AGAGATGACGAACCTGCTCCGAATC 962

	RESULT	2					18-FEB-2000
	LOCUS	AR069366/c					
	DEFINITION	AR069366	4411 bp	DNA	PAT		
	ACCESSION	Sequence 5 from patent US 5891637.					
	VERSION	AR069366					
	KEYWORDS	AR069366.1 GI:7220254					
	SOURCE	. Unknown.					
	ORGANISM	Unknown.					
	REFERENCE	Unclassified.					
	AUTHORS	1 (bases 1 to 4411)					
	TITLE	Ruppert,S.J.W.					
	JOURNAL	Construction of full length cDNA libraries					
	FEATURES	Patent: US 5891637-A 5 06-Apr-1999;					
		Location/Qualifiers					
	source	1..4411					
	BASE COUNT	/organism='unknown'					
	ORIGIN	1065 a 1156 c 1124 g 1066 t					

RESULT		3	PAT :	9-8-FEB-2000
LOCUS	AR069362			
DEFINITION	AR069362	5443 bp	DNA	
ACCESSION	Sequence 1 from patent US 5891637.			
VERSION	AR069362			
KEYWORDS	AR069362.1 GI:7220250			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 5443)			
TITLE	Ruppert,S.J.W.			
JOURNAL	Construction of full length cDNA libraries			
FEATURES	Patent: US 5891637-A 1 06-Apr-1999;			
source	location/Qualifiers 1..5443 /organism="unknown"			
BASE COUNT	1268 a 1441 c 1446 g 1288 t			
ORIGIN				
Query Match	8.1%:	Score 64.8:	DB 9:	Length 5443;
Best Local Similarity	97.1%:	Pred. No.8.4e-07;		
Matches 66; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
OY	737 ctgtctccgaattcgagcgcgctcgacaagaacttgacgccacctggagcacaccaccaaac	796		
Dd	5240 CGAGTCCGAATTTCGACGTCTCGACAAGCTTTGGGCCGCCTACTCGAAGCACCAACCACACACC	5299		
OY	797 accactga 804			
Dd	5300 ACCACTGA 5307			
RESULT		4		

AR069364	LOCUS	AR069364	5616 bp	DNA	PAT	18-FEB-2000
DEFINITION	Sequence 3 from patent US 5891637.					
ACCESSION	AR069364					
VERSION	AR069364.1	GI:7220252				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 5616)					
AUTHORS	Ruppert,S.J.W.					
TITLE	Construction of full length cDNA libraries					
JOURNAL	Patent: US 5891637-A 3 06-Apr-1999;					
FEATURES	location/Qualifiers					
source	1..5616					
	"/organism="unknown"					
BASE COUNT	1306 a 1478 c 1492 g 1340 t					
ORIGIN						

LOCUS	SEQUENCE	FROM PATENT	US	5891637	PAT	18-FEB-2000
LOCUS	AR054309	2735 bp	DNA			29-SEP-1999
DEFINITION	Sequence 1	from patent	US	5837240		
ACCESSION	AR054309					
VERSION	AR054309.1	GI:5979886				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 2735)					
	Iee,C.K., Monath,T.P., Ackerman,S.K., Thomas,W.D., Soman,G.,					
	Kleanthous,H., Welzlin,R.A., Pappo,J., Ernack,T., Guirakhoo,F.,					
	Bagatel,H., and Sussman,I.					
TITLE	Multimeric, recombinant urease vaccine					
JOURNAL	Patent: US 5837240-A 1 17-NOV-1998;					
FEATURES	Location/Qualifiers					
source	1..2735					
BASE COUNT	875 a	562 c	635 g	663 t		
ORIGIN	/organism="unknown"					
Query Match	7.6%;	Score	61.4;	DB	9;	Length 2735;
Best Local Similarity	98.4%;	Pred. No.	7e+06;			
Matches	62;	Conservative	0;	Mismatches	0;	Indels 0;
						Gaps 0;
QY	742 ccgaattcgagctccgtcgcgaagaagcttgccgcgcgaacccgcagcaccacacacacacac	801				
	1					
Db	2562 CTGAATTCGACCTCGTCGCAACAAGCTTGCGCGCGCAGCTCGAGCACCACACACACAC	2621				
QY	802 tga	804				
Db	2622 TGA	2624				
RESULT	6					
AR059367	6363 bp	DNA				
LOCUS	AR059367	6363 bp	DNA			
DEFINITION	Sequence 6 from patent	US	5891637			
PAT						

ACCESSION AR069367
VERSION AR069367.1 GI:7220255
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6363)
AUTHORS Rupert,S.J.W.
TITLE Construction of full length cDNA libraries
JOURNAL Patent: US 5891637-A 6 06-APR-1999;
FEATURES
source
BASE COUNT 1576 a 1701 c 1645 g 1441 t
ORIGIN

Query Match 6.8%; Score 55; DB 9; Length 6363;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 gagctccgtcgaagaactgtgcgcgaactcgcagcaccaccaccaccactga 804
|||||
Db 4656 GAGCTCCGTCGACACGCTTGCGGCCGCTACTCGACACACACACACCTGA 4710

RESULT 7
BSPBGA1G
LOCUS BSPBGA1G 1416 bp DNA BCT 07-NOV-1994
DEFINITION Bacillus sp. bgaa gene for lichenase.
ACCESSION 212151
VERSION 212151.1 GI:296931
KEYWORDS bgaa gene; lichenase.
SOURCE Bacillus sp.
ORGANISM Bacillus sp.
REFERENCE 1 (bases 1 to 1416)
AUTHORS Taberner,C., Coll,P.M., Fernandez-Abalos,J.M., Perez,P. and Santamaria,R.I.
TITLE Cloning and DNA sequencing of bgaa, a gene encoding an endo-beta-1,3-1,4-glucanase, from an alkalophilic Bacillus strain (N137)
JOURNAL Appl. Environ. Microbiol. 60 (4), 1213-1220 (1994)
MEDLINE 94288605
REFERENCE 2 (bases 1 to 1416)
AUTHORS Santamaria,R.I.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1992) Ramon I. Santamaria, Instituto de Microbiologia Bioquimica, C.S.I.C./ Universidad de Salamanca, Plaza de la Merced S/N, Salamanca, Salamanca, 37008, Spain
FEATURES
source
1. 1416
/organism="Bacillus sp."
/strain="N137"
/db_xref="taxon:1409"
/clone="10"
199..204
222..227
235..387
376..379
388..1218
/gene="bgaa1"
388..1218
/gene="bgaa1"
/codon_start=1
/transl_table=11
/product="lichenase"
/protein_id="CAA78135.1"
/db_xref="GI:296931"
/db_xref="SPTREMBL:O45648"
/translation="MKKLVLFSCMLLFGSLISGLVSPQVAEAEREIGTAPEVET
ESYDSERSKAGVSNQGMFNATWYPEQVTFPSNGMKLIQIEKEDNETASPKAGELR

TNDFFHYGLFEVSMKPAKSTGVSSFEYTGPMWENDEPWEIDIEFLKDTTKIQFN
YFTNGVGGNEHYHEIDPAADENFYAEPMREESI.RMFVNGELVTAETENIPQTOKI
MMNLMPGICVDGWTGRFNGEDIPVYTDWVRYTFLEELGCTENKNNKKCKRTKVK
"

sig_peptide 388..480
mat_peptide 481..1215
/gene="bgaa1"
/product="lichenase"

BASE COUNT 509 a 181 c 276 g 450 t
ORIGIN

Query Match 6.2%; Score 50; DB 2; Length 1416;
Best Local Similarity 51.5%; Pred. No. 0.0084;
Matches 172; Conservative 0; Mismatches 150; Indels 12; Gaps 2;

QY 25 ggtccgaactctacacgttagaagaagttcagttacgttagaagttgaagccgtatgaag 84
|||||
Db 679 GCTGGAGAACTTCGACAAACGATTTTATCATTTACGTGGGTTGTGAAGTGAATGAAA 738

QY 85 atggcagccgcgacgcggaagaagttcagttcctctaccagaatggtccgaatc 144
|||||
Db 739 CCTGCAAGTCAACGGGAACAGCTCTTCAATTTTACCTATACCTGACCTTGGAATGG 798

QY 145 gccgatggaagcccggttagaagttgatattgaagttctgcgcaagaatccgagcagt 204
|||||
Db 799 GAGAAATGAT---CCATGGGATGGAATTTGATATGCAATTTTATAGTAAGTACTATAA 855

QY 205 ttccagttccacacatcattaccggttagagccgcgcacaaagaatagcgaagaacacat 264
|||||
Db 856 ATACAAATTAATTAATTTTACAAACGGAGTAGCGGAAATGACGATP-----ACCAT 906

QY 265 gctgttagcccgccgcgcgcatcaggtttccacacctagcgttcgaatgactccgaat 324
|||||
Db 907 GAATTAGGATTTATGTCAGCAGATGATTTTAATACATGCTTTTGAAGTGAACACGAA 966

QY 325 taagtcgcgtgactgttagcaggttcaggaagtc 358
|||||
Db 967 TCTATTCGTTGGTTTGTAAATGCTGAATTCGTTTC 1000

RESULT 8
SYHA78M
LOCUS SYHA78M 1234 bp DNA SYN 04-APR-2000
DEFINITION Synthetic B.macerans/B.amyloliquefaciens hybrid gene for endo-1,3-1,4-beta-glucanase.
ACCESSION 225882
VERSION 225882.1 GI:398202
KEYWORDS beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE Synthetic construct.
ORGANISM Artificial sequence.
REFERENCE 1 (bases 1 to 777)
AUTHORS Hofemeister,J., Kurtz,A., Borris,R. and Knowles,J.
TITLE The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis
JOURNAL Gene 49 (2), 177-187 (1986)
MEDLINE 87192007
REFERENCE 2 (bases 396 to 852)
AUTHORS Borris,R., Buettner,K. and Meentaseae,J.P.
TITLE Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans: homologues to other beta-glucanases
JOURNAL Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
MEDLINE 91109712
REFERENCE 3 (bases 1 to 1234)
AUTHORS Olsen,O., Borris,R., Simon,O. and Thomsen,K.K.
TITLE Hybrid Bacillus (1-3,1-4)-beta-glucanases: engineering thermostable enzymes by construction of hybrid genes
JOURNAL Mol. Gen. Genet. 225 (2), 177-185 (1991)
MEDLINE 91172113
REMARK (sites)
4 (bases 1 to 1234)

AUTHORS	Poltz,O., Simon,O., Olsen,O. and Borris,R.
TITLE	Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases
JOURNAL	Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE	94009045
REFERENCE	5 (bases 1 to 1234)
AUTHORS	Poltz,O.
TITLE	Direct Submission
JOURNAL	Submitted (26-AUG-1993) Poltitz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
FEATURES	
source	1..1234
	/organism="synthetic construct"
	/db_xref="taxon:32630"
source	/focus
	1..1234
	/organism="Bacillus amyloliquefaciens"
	/db_xref="taxon:1390"
source	1..777
	/organism="Bacillus amyloliquefaciens"
	/db_xref="taxon:1390"
	1..1234
	/organism="Paenibacillus macerans"
	/db_xref="taxon:44252"
sig_peptide	469..543
CDS	469..1188
	/function="endo-1,3-1,4-beta-glucanase"
	[note]"Hybrid between bgla (Pos. 1 to 777) and bglm (pos.in hybrid 778 to 1234, pos. original 396 to 852)"
	/codon_start=1
	/label_table=1
	/label_HA78M
	/evidence="experimental"
	/product="hybrid-endo-1,3-1,4-beta-glucanase"
	/protein_id="CAAB1101.1"
	/db_xref="GI:398203"
	/translation="MKRVLLILVTGLFMSLCGITSVSQAOTGGSFPEFPNSYNSLQWQKADYSODGNMFNCTIRANNVSMTSLGMERIALDITSPSYNKFDGENRSVOYGLEYEVRAKRPAKNKTIVSEFFPYTPGAHGTOMBEDIDLEFLGDPTTKVFONYNTNGVGSHKVISTEIGFSAPKHFTAFDMOPGYIKWYVDVLKKTTATANIPIPSFGKIIMNLINMGTVDDMLGSYNGANPLVAEYMVKYTSSN"
source	778..1234
	/organism="paenibacillus macerans"
	/db_xref="taxon:44252"
BASE COUNT	373 A 224 C 281 G 356 T
ORIGIN	
Query Match	6.2%; Score 49.6; DB 56; Length 1234;
Best Local Similarity	51.8%; Pred.No. 0.011;
Matches 174,	Conservative 0; Mismatches 144; Indels 18; Gaps 2;
OY	13 aaagatttagcgggtgcggaactctaacgcgtttagaagaagtctcgtaagtgtaa 72
DB	712 AACAA GTTGACTCGGGGGAAAAACCGTCGGTCCAACAATAAGCCTTAAGCATTTATGA 771
OY	73 gcccgatgaagatbgccgcgcgatccggaagaaagtcaggcttccatgttctcaagaat 132
DB	772 CTCCAA ATGAAGCCACCAGAAAATAATCACAGAATTGTCTCATCTTTTCACGTA----- 824
OY	133 ggttcgaataatgcccgatgaagaagcccctgggtlagaagtgbatatgaagttccgcaag 192
DB	825 --TACAG ACCTGCATCATGCCAACACATAGGGATGAATAATGCAATTTCTTAGGAAAA 882
OY	193 ataccggcagttccaatccaacaaccttatcteccggltaaggccggcacaaaagactgc 252
DB	883 GACAC GAACAAAAGTCCAGTTTAACATTATATATACCAATGGGGTTGGGCGTCAT----- 933
OY	253 gaaagacaecatcygtttaagccccgcgcgcgctaagagcttccacacctaatgyctcgaa 312
DB	934 GAAAG AGTTATCTCTCTTGCGCTTTTATGCAATGAAGGGCTTCACATACCTATGCTTTGCAT 993
OY	313 tggactccgaattacgtccgtcgtgaactgltgaagyt 348

```

Db      994   TGGCAGCCAGGCTAATTAATGCATTCGTACAGCGT    1029
|||||  |||  |||  |  |||  |||  |||||
RESULT  9
AX026763          AX026763       98 bp     DNA             PAT        16-SEP-2000
LOCUS           Sequence 7 from Patent WO0040712.
DEFINITION      AX026763
ACCESSION       AX026763
VERSION         AX026763.1 GI:10187902
KEYWORDS
SOURCE          . synthetic construct.
ORGANISM        artificial construct.
REFERENCE       1 (bases 1 to 98)
AUTHORS        Tawfik,D. and Griffiths,A.
TITLE          Optical sorting method
JOURNAL        Patent: WO 0040712-A 7 13-JUL-2000;
FEATURES       MEDICAL RES COUNCIL (GB); TAMFIK DAN (GB); GRIFFITHS ANDREW (GB)
source         Location/Qualifiers
               1..98
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="per-23d FLAG polylinker"
               3..>29
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=1
                /protein_id="CAC09112.1"
                /db_xref="GI:10187903"
                /translation="MDYKDDDDK"
BASE COUNT     30 a            26 c            26 g            16 t
ORIGIN
Query Match    6.0%: Score 48.6; DB 9; Length 98;
Best Local Similarity 92.7%; Pred. No. 0.022;
Matches 51; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy      731   acgaacctgcctccgaatcgcagctcgctcgcaagaacttgcgacgactggaca 785
||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      44   AGGTACCGGATTCGAATTCGAGCTCCGTCGACACACTTGCGGCCGCACTCGACGA 98
RESULT  10
AX000165          AX000165       927 bp     DNA             PAT        10-MAR-2000
LOCUS           Sequence 2 from Patent WO9906573.
DEFINITION      AX000165
ACCESSION       AX000165
VERSION         AX000165.1 GI:7240678
KEYWORDS
SOURCE          unidentified.
ORGANISM        unclassified.
REFERENCE       1 (bases 1 to 927)
AUTHORS        Maurer,K. and Hillen,W.
TITLE          NEW BETA-GLUCANASE FROM A BACILLUS
JOURNAL        Patent: WO 9906573-A 2 11-FEB-1999;
FEATURES       MAURER KARL HEINZ (DE); HENKEL KRAA (DE)
source         Location/Qualifiers
               1..927
                /organism="unidentified"
                /strain="BACILLUS ALKALOPHILUS DSM 9956"
                /db_xref="taxon:32644"
BASE COUNT     343 a            150 c            218 g            216 t
ORIGIN
Query Match    5.9%: Score 47.4; DB 9; Length 927;
Best Local Similarity 50.7%; Pred. No. 0.043;
Matches 173; Conservative 0; Mismatches 156; Indels 12; Gaps 2;
Oy      16   gatttaacgctgtccgaactctaacgcttaagaagaagttaacgctaagttgaagcc 75

```

```

Db      286 GAATATAGCTGGCGAATTAAGACGAATCAGTTTATATCAATACGGGTTGTTGAAGTC 345
Qy      76  cgtatgaagaatgcaagccgcatcggaagacagtaagttccatgttccctccgaagatg 135
Db      346 AATATGAGCCAGCAGGAATACAGGACCGCTCTTCTACTCTTTACATATACGGGTCGA 405
Qy      136 tccgaataatcgccagatggaagccctggtagaagtgatataatgaatcccgcaagaat 135
Db      406 TGGGATTGGGATATATG---ATCCTTGGGATGAATGATATGATGCTTCCCTTGAAGGAT 462
Qy      196 cgggagcagttccagtcacatcaatcaacgcggaagcgcgcaacaagaactagcgaa 255
Db      463 ACACAAGAGCTCCAAATTAATTAATTTACTTAA-----CCGAGTAGCAACAAATGAA 513
Qy      256 aagcaatgctgttagcccgccgcatcagagcttccacacccatcagttccgaatg 315
Db      514 CATTACACAGCAATTAAGGCTTCGATGCAATCAATCTTTTAATACGTAATGCTTTGAATGG 573
Qy      316 actcgaatcagtcgcgctgagctgttgaagcgtcaggaagt 356
Db      574 AGACGAGATCATTAATTAATGTTGTAAGTAACGGAATTAAGT 614

```

```

RESULT 11
ASTCELL90
LOCUS      306 bp      DNA      SYN      08-NOV-1996
DEFINITION Artificial sequence DNA for T-cell receptor, toolbox 90.
ACCESSION Y09275
VERSION    Y09275.1 GI:1666290
KEYWORDS   T-cell receptor; toolbox 90.
SOURCE     synthetic construct.
ORGANISM   artificial sequence.
REFERENCE  1 (bases 1 to 306)
AUTHORS    Schluesener, H.J.
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 306)
AUTHORS    Schluesener, H.J.
TITLE      Direct Submission
JOURNAL    Submitted (21-OCT-1996) H.J. Schluesener, Inst.f. Hirnforschung,
          Calver Str. 3, D-77076 Tuebingen, FRG
FEATURES   source
            location/Qualifiers
            1..306
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="toolbox 90"
            /codon_start=1
            /transl_table=11
            /product="T cell receptor peptide"
            /protein_id="CAA70470.1"
            /db_xref="GI:1666291"
            /translation="MASMTGGCGQMGKGFPAADPGTIOQLFMSFNNQSESLVELDM
            GHGLRLIHVSYSGSPFENGVDMDHGLRLIHVSIDYNSKLEKLAALHNNHHH"
BASE COUNT 65 a      103 c      86 g      52 t
ORIGIN

```

```

Query Match      5.7%: Score 45.8; DB 56; Length 306;
Best Local Similarity 95.9%: Pred. No. 0.12;
Matches 47; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy      756 cgtgacaagaatgctggcgccgcaatcgagacacacacacacacacactga 804
Db      258 CCTCGAGAAGCTTGGCGCGCACTCGACCAACCAACCAACCAACCACTGA 306
RESULT 12
SYNHAOM      1228 bp      DNA      SYN      04-APR-2000
LOCUS        Synthetic B. macerans/B. amylioliquefaciens hybrid gene for
DEFINITION   endo-1,3-1,4-beta-glucanase.

```

```

ACCESSION 225873
VERSION   225873.1 GI:398185
KEYWORDS  beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE    synthetic construct.
ORGANISM  artificial sequence.
REFERENCE  1 (bases 1 to 543)
AUTHORS    Hofemeister, J., Kurtz, A., Borris, R. and Knowles, J.
TITLE      The beta-glucanase gene from Bacillus amylioliquefaciens shows
          extensive homology with that of Bacillus subtilis
JOURNAL    Gene 49 (2), 177-187 (1986)
MEDLINE    87192007
REFERENCE  2 (bases 168 to 852)
AUTHORS    Borris, R., Buettner, K. and Maentzsaellae, P.
TITLE      Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
          homologues to other beta-glucanases
JOURNAL    Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
MEDLINE    91109712
REFERENCE  3 (bases 1 to 1228)
AUTHORS    Polltz, O., Simon, O., Olsen, O. and Borris, R.
TITLE      Determinants for the enhanced thermostability of hybrid
          (1-3,1-4)-beta-glucanases
JOURNAL    Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE    94009045
REFERENCE  4 (bases 1 to 1228)
AUTHORS    Polltz, O., Simon, O., Olsen, O. and Borris, R.
TITLE      Determinants for the enhanced thermostability of hybrid
          (1-3,1-4)-beta-glucanases
JOURNAL    Eur. J. Biochem. (1993) In press
REMARK     (sites)
AUTHORS    5 (bases 1 to 1228)
JOURNAL    Polltz, O.
TITLE      Direct Submission
JOURNAL    Submitted (26-AUG-1993) Polltz O., Humboldt University of Berlin,
          Biology, Marschauer Str. 43, BERLIN, Germany, D-10243
FEATURES   source
            location/Qualifiers
            1..1228
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /focus
            1..1228
            /organism="Bacillus amylioliquefaciens"
            /db_xref="taxon:1390"
            1..543
            /organism="Bacillus amylioliquefaciens"
            /db_xref="taxon:1390"
            1..1228
            /organism="Paenibacillus macerans"
            /db_xref="taxon:44252"
            469..543
            /function="endo-1,3-1,4-beta-glucanase"
            /note="hybrid between bgia (pos. 1 to 543) and bgim
            (pos. in hybrid 544 to 1228, pos. original 168 to 852)"
            /codon_start=1
            /transl_table=11
            /label="HAOM"
            /evidence="experimental"
            /product="hybrid endo-1,3-1,4-beta-glucanase"
            /protein_id="CAA81092.1"
            /db_xref="GI:398186"
            /translation="MKRVLLILVGLFMSLCITSSVAGSVFMEPLSYFNPSTWEKA
            DYSNMGVNCNTPRANVNFNNDCKIKGLTSSVANKFDCAEVRSNTYNGVGLKYSIG
            KPAKNTGIVSSFFETTGPAHGTOMDEIDIEFGKDTTVQGNVTYTNCGVGEKYSISG
            FDASKGFHTVAFDQPGYIKWYVDGLVLTHTATNPSTPGKIMNLMNGTVDWGLGS
            YGNAPLVAEYDWMYKTSN"
            544..1228
            /organism="Paenibacillus macerans"
            /db_xref="taxon:44252"
BASE COUNT 371 a      220 c      290 g      347 t
ORIGIN

```

Query Match	5.5%	Score 44.6	DB 56	Length 1228
Best Local Similarity	52.9%	Pred. No. 0.24		
Matches 154	Conservative	0	Mismatches 119	Indels 18
			Gaps	2
QY	58	tacgaaagatttgaagcccgatagagatlggaagcgcatcgggaacagctcagttccatg	117	
DB	751	TACGGCCGTACAGAGGTGAGTGTGAAGCCAGCCAAAATACAGGAATGTCTCATTCCTT	810	
QY	118	ttccctcacccaagatggttcggaatcgcgcgaagaaagccctgggtlagaatgtgataat	177	
DB	811	TTACAGTA-----TACAGACCTCTCATGCGACACACATGGGATGAAATAGATATAC	861	
QY	178	gaagttctcgcgaagaatcgcggcgagtttcagttccagttcaacaatcatatccggttaagccgcgc	237	
DB	862	GAATTTCTTAGGAAAAGACACGACAAAAGTCCAGTTTAATACTATTATACCAATGGGGTTGGC	921	
QY	238	gcacaaagaactagagaaaagcacatgctgtttagccccgcgcgcgcatcaggctttccac	297	
DB	922	GGTCTAT-----GAAAAGGTTATCTCTCTGGCTTTGATGACATCAAGGGCTTCCTCAT	972	
QY	298	aactcaggtctcgaatgcgaatccgaattacgctgcgagctgttgagcgt	348	
DB	973	ACCTATGCTTTTCGATTGGCAGCCAGCGATATTAAATGATGTATGACGCT	1023	

RESULT	13
SYHA12Y13-LOCUS	
DEFINITION	SYHA12Y13 1234 bp DNA SYN 04-APR-2000 Synthetic B.macerans/B.amyloliquefaciens hybrid gene for endo-1,3-l,4-beta-glucanase.
ACCESSION	225876
VERSION	225876.1 GI:398190
KEYWORDS	beta-glucanase;.endo-1,3-l,4-beta-glucanase. synthetic construct. synthetic construct. artificial sequence.
SOURCE	1 (bases 1 to 579) Hofemeister,J., Kurtz,A., Borris,R. and Knowles,V. The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis
ORGANISM	Gene 49 (2), 177-187 (1986)
TITLE	87192007 2 (bases 198 to 852) Borris,R., Buettner,K. and Maentzsaefae,P. Structure of the beta-1,3-l,4-gluacanase gene of Bacillus macerans homologes to other beta-glucanases Mol. Gen. genet. 222 (2-3), 278-283 (1990)
JOURNAL	91109712 3 (bases 1 to 1234) Polltz,O., Simon,O., Olsen,O. and Borris,R. Determinants for the enhanced thermostability of hybrid (1-3,l-4)-beta-glucanases Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE	9409045 4 (bases 1 to 1234) Polltz,O., Simon,O., Olsen,O. and Borris,R. Determinants for the enhanced thermostability of hybrid (1-3,l-4)-beta-glucanases Eur. J. Biochem. (1993) In press
REFERENCE	(sites) 5 (bases 1 to 1234) Polltz,O. Direct Submission Submitted (26-AUG-1993) Polltz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
AUTHORS	Location/Qualifiers . .1234 /organism="synthetic construct" /db_xref="taxon:32630" /focus . .1234 /organism="Bacillus amyloliquefaciens" /db_xref="taxon:1390"
JOURNAL	
TITLE	
FEATURES	
source	
Source	

```

source
1. 579
/organism="Bacillus amyloliquefaciens"
/db_xref="taxon:1390"
1. 1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
469. 543
469. 1188
/function="endo-1,3-1,4-beta-glucanase"
/note="Hybrid between bgla (Pos. 1 to 579) and bg1m
(Pos. In hybrid 580 to 1234, Pos. Original 198 to 852)
Deletion of Y13 of HA12-M"
/codon_start=1
/transl_table=11
/label=HA12M
/evidence=experimental
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CAAB1095.1"
/db_xref="GI:398191"
/translacion="MKRVLILVTGLFMSLCGTSVSAQTGSGFEFPNSPSTWE
KADYSGVGENCTIRANWNVFTNDGLKLGITSSAVNKFDCAEYRSTNIYGLYEV
SMRKAKTNYHYSSEFTGYTGAGHGTOMOEIDIEFGKDTTVQFNVYVNGGGEKYS
LGPEASKGFTHYAEDMQPGYIKWYVDVLKHTATANIPIPTPKIMNMLNNGTGVDML
GSGINAGPLTAEIYMWVYTSN"
580. 1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"

```

Query Match	5.5%	Score 44.6	DB 56	Length 1244
Best Local Similarity	52.9%	Pred. No. 0.24	Mismatches 119	Indels 18
Matches 154	Conservative	0	Gaps	2
Qy 58	tacgtaagtttcgaagccgcgatgaagatgycagccgacatcggagacatcagttccatg	117		
Db 757	TACGGCCCTGTACGAGGTGATGAAGCCACGCAAAAATACAGGAATGTCTCATCTTT	816		
Qy 118	ttcccttcaccgaataggttcggaatcgccgatggaagccctggtgtgaagtgaattc	177		
Db 817	TTACACGTA-----TACAGGACCTCTCATGGCACACAAATGGATGAATGAATATC	867		
Qy 178	gaagttcccggaagatcgcggcagttccagttccacatcatcagcgaagccgcg	237		
Db 868	GAATTTTAGGAAAGACACGACGAAAAAGTCCAGTTTAACTATTACCAATGGGGTTGC	927		
Qy 238	gcacaaagactagcgaagaagaccatctgtttagcccgccgcgcatcaggtttcac	297		
Db 928	GCTCAT-----GAAAGAGTTATCTCTCTTGCGTTGATGCAATCAAGCGCTTCAT	978		
Qy 298	acctcaggttcgaatgaatccgaatcagtcgcgcgtgactgttgaaggt	348		
Db 979	ACCTATGCTTTGATTTGGACGACGAGGTATATTAAATGTTATGTATGACGGT	1029		
RESULT 14				
SYHA16M	LOCUS	SYN	04-APR-2000	
DEFINITION	SYNTHETIC B. macerans/B. amyloliquefaciens hybrid gene for endo-1,3-1,4-beta-glucanase.			
ACCESSION	Z25878			
KEYWORDS	Z25878.1 GT:398194			
TERMS	beta-glucanase; endo-1,3-1,4-beta-glucanase.			
SOURCE	synthetic construct.			
ORGANISM	synthetic construct			
REFERENCE	artificial sequence.			
AUTHORS	1 (bases 1 to 591)			
TITLE	Hotemister, J., Kurtz, A., Borriis, R. and Knowles, J.			
JOURNAL	The beta-glucanase gene from <i>Bacillus amyloliquefaciens</i> shows extensive homology with that of <i>Bacillus subtilis</i>			
MEDLINE	gene 49 (2), 177-187 (1986)			
	87192007			

Oy	178	gaagcttcgcgcaagaatccggcgacgtttccagcttccaacatacatcaggtaagccgcg	237
Dd	868	GAAATTCTAGGMAAAGCACGACAAAAGTCCACTTTAACTATTAATRCCAATGGGTTCGC	927
Oy	238	gcacaaagaactagcgaaagaccatcgtcgttagccccgcgcgatcaggctttccac	297
Dd	928	GGTCACT-----GAAGAAGTATATCTCCTTGCGTTTGATGATCAACAAGGCGTTTCAT	978
Oy	298	aacctacgcttcgaaibygactccggaattacgttcgcgttgacigcgttgacggt	348
Dd	979	ACCTATGCTTTGATGATGGCAGCAGGGGTATATTAAATGGTATGTAGACGCT	1029
RESULT	15		
LOCUS	SYHA2M	1234 bp	DNA SYN 04-APR-2000
DEFINITION	Synthetic B.macerans/B.amyloliquefaciens hybrid gene for endo-1,3-1,4-beta-glucanase.		
ACCESSION	225879		
VERSION	225879.1	GI:398196	
KEYWORDS	beta-glucanase; endo-1,3-1,4-beta-glucanase.		
SOURCE	synthetic construct.		
ORGANISM	synthetic construct.		
REFERENCE	artificial sequence.		
AUTHORS	1 (bases 1 to 549)		
TITLE	Hofmeister,J., Kurtz,A., Borrijs,R. and Knowles,J. The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis Gene 49 (2), 177-187 (1986)		
JOURNAL	2 (bases 168 to 852)		
MEDLINE	Borrijs,R., Buettner,K. and Maentzelae,P.		
REFERENCE	Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans: homologies to other beta-glucanases		
AUTHORS	Mol. Gen. genet. 222 (2-3), 278-283 (1990)		
TITLE	91109712		
JOURNAL	3 (bases 1 to 1234)		
MEDLINE	Poltz,O., Simon,O., Olsen,O. and Borrijs,R.		
REFERENCE	Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases		
AUTHORS	Eur. J. Biochem. 216 (3), 829-834 (1993)		
TITLE	94009045		
JOURNAL	4 (bases 1 to 1234)		
MEDLINE	Poltz,O., Simon,O., Olsen,O. and Borrijs,R.		
REFERENCE	Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases		
AUTHORS	Eur. J. Biochem. (1993) In press		
JOURNAL	(sites)		
REMARK	5 (bases 1 to 1234)		
REFERENCE	Poltz,O.		
AUTHORS	Direct Submission		
TITLE	Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243		
JOURNAL	Location/Qualifiers		
FEATURES	1..1234		
source	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	/focus		
source	1..1234		
	/organism="Bacillus amyloliquefaciens"		
	/db_xref="taxon:1390"		
source	1..549		
	/organism="Bacillus amyloliquefaciens"		
	/db_xref="taxon:1390"		
source	1..1234		
	/organism="Paenibacillus macerans"		
	/db_xref="taxon:44252"		
sig_peptide	469..543		
CDS	469..1188		
	/function="endo-1,3-1,4-beta-glucanase"		
	/note="Hybrid between bglA (Pos. 1 to 549) and bglm (pos.in hybrid 550 to 1334, Pos. Original 168 to 852)"		

```

/codon_start=1
/transl_table=11
/label=HA2M
/evidence=experimental
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CA81098.1"
/db_xref="GI:3981197"
/translation="MKRVLLIYLTLGLFMSLCITSSVSNQSGSVHMEPLSLFNSNTHME
KADQSSKGVGFNCTPYTANANNVFNDEKILGLITSSAYNKFQDAEYRSTNITGLYLEV
SMKRAKNGVIVSSEFTYTPAHQTOWDELDIEFLKDTTKQVQFNNTYGVGSGHRYLV
LGFASVAGCEFTYAFMDQPGYIKWYVDGVLTHTATANIPISTPGKIMMNLNCTGVDWML
GSYNGARPLVLAEDWVKRYNSN"
550..1234

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:16:57 ; Search time 157.43 Seconds
(without alignments)
3206.713 Million cell updates/sec

Title: US-09-654-652a-5

Perfect score: 804
1 atcgttagcgcacaagattt.....accacaccaccacacactga 804

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /cgnl_9/gcgdata/geneseq/NA1980.DAT:*
- 2: /cgnl_9/gcgdata/geneseq/NA1981.DAT:*
- 3: /cgnl_9/gcgdata/geneseq/NA1982.DAT:*
- 4: /cgnl_9/gcgdata/geneseq/NA1983.DAT:*
- 5: /cgnl_9/gcgdata/geneseq/NA1984.DAT:*
- 6: /cgnl_9/gcgdata/geneseq/NA1985.DAT:*
- 7: /cgnl_9/gcgdata/geneseq/NA1986.DAT:*
- 8: /cgnl_9/gcgdata/geneseq/NA1987.DAT:*
- 9: /cgnl_9/gcgdata/geneseq/NA1988.DAT:*
- 10: /cgnl_9/gcgdata/geneseq/NA1989.DAT:*
- 11: /cgnl_9/gcgdata/geneseq/NA1990.DAT:*
- 12: /cgnl_9/gcgdata/geneseq/NA1991.DAT:*
- 13: /cgnl_9/gcgdata/geneseq/NA1992.DAT:*
- 14: /cgnl_9/gcgdata/geneseq/NA1993.DAT:*
- 15: /cgnl_9/gcgdata/geneseq/NA1994.DAT:*
- 16: /cgnl_9/gcgdata/geneseq/NA1995.DAT:*
- 17: /cgnl_9/gcgdata/geneseq/NA1996.DAT:*
- 18: /cgnl_9/gcgdata/geneseq/NA1997.DAT:*
- 19: /cgnl_9/gcgdata/geneseq/NA1998.DAT:*
- 20: /cgnl_9/gcgdata/geneseq/NA1999.DAT:*
- 21: /cgnl_9/gcgdata/geneseq/NA2000.DAT:*
- 22: /cgnl_9/gcgdata/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	65.6	8.2	5248 16	AAAT08489
2	65.6	8.2	5248 17	AAAT1898
3	64.8	8.1	5616 20	AAAX2808
4	64.8	8.1	5616 20	AAAX2808
5	61.4	7.6	2735 17	AAAT4351
6	61.4	7.6	2007 20	AAAX80369
7	61.4	7.6	2007 20	AAAX80370
8	61.4	7.6	2007 20	AAAX80366
9	61.4	7.6	2007 20	AAAX80367
10	61.4	7.6	2007 20	AAAX80368
11	55.5	6.8	6363 20	AAAX8091

12	50.8	6.3	59	22	AAAT74635
13	47.4	5.9	927	20	AAAX02912
14	44.6	5.5	783	11	AAAO05167
15	44.6	5.5	850	11	AAAO03519
16	42.6	5.3	586	21	AAAG1166
17	42.4	5.3	423	20	AAAT19396
18	42.4	5.3	423	21	AAAO5832
19	42.2	5.2	1240	11	AAAO19184
20	42.2	5.2	357	21	AAAO5832
21	42.2	5.2	1186	20	AAAG1167
22	42.2	5.2	1247	20	AAAX29725
23	42.2	5.2	6619	20	AAAG3789
24	41.6	5.2	550	21	AAAT1987
25	38.8	4.8	1815	19	AAAT39341
26	38.8	4.8	1815	19	AAAV1538
27	37.6	4.7	1166	20	AAAT19440
28	37.6	4.7	1166	20	AAAT19228
29	37.4	4.7	607	21	AAAG4807
30	37.2	4.6	241	21	AAAG4064
31	35.6	4.4	1029	22	AAAF60750
32	35.2	4.4	445	21	AAAT1502
33	35.2	4.4	1195	20	AAAT19453
34	35.2	4.4	1195	20	AAAT19241
35	34.6	4.3	556	21	AAAF09632
36	34.2	4.3	948	16	AAAG0748
37	34.2	4.3	948	17	AAAT27687
38	34.2	4.3	948	18	AAAT70349
39	34.2	4.3	948	18	AAAT76649
40	34.2	4.3	948	19	AAAV5799
41	34.2	4.3	948	19	AAAV5873
42	34.2	4.3	948	20	AAAV63425
43	34.2	4.3	963	18	AAAT76626
44	34.2	4.3	963	19	AAAV5831
45	34.2	4.3	963	19	AAAV53941

ALIGNMENTS

RESULT 1	
AAAT08489	AAAT08489 standard; DNA; 5248 BP.
AC	AAAT08489;
XX	
DT	15-MAR-1996 (first entry)
XX	
DE	Plasmid PET TRC SO5/NI containing 561.1 scFv.
XX	
KW	Complement C5; haemolysis; kidney; glomerulonephritis;
KW	monoclonal antibody; antiinflammatory; antibody engineering;
KW	humanised antibody; complementarity determining region; CDR;
KW	single chain antibody; scFv; vector; PET TRC SO5/NI;
RW	Escherichia coli; ds.
XX	
OS	Synthetic.
XX	
FN	W09529697-A1.
XX	
PD	09-NOV-1995.
XX	
PF	01-MAY-1995; 95WO-US05688.
XX	
PR	02-MAY-1994; 94US-0236208.
XX	
PA	(ALEX-) ALEXION PHARM INC.
XX	
PI	Evans MJ, Matlis L, Mueller EE, Nye SH, Rollins S;
PI	Rother RP, Springhorn J P, Squinto SP, Thomas TC;
XX	Wang Y, Wilkins JA;
DR	WPI; 1995-392923/50.
XX	

Cloning vector pRK
B. alkalophilus de
Heat-stable endo-b
Thermostable beta-
SEN virus genome f
M. tuberculosis an
Beta-glucanase-H1
SEN virus genome f
Insert from pET28C
Insert from pET28C
Plasmid pKMV2768 e
Human VEGF-X CUB-1
Human papillomavir
Human papillomavir
M. tuberculosis an
M. tuberculosis re
C. trachomatis clo
H. contortus clone
Pseudomonas sp ABC
Neocarzinostatin a
M. tuberculosis re
Fusarium venenatum
5' Nuclease from T
Mutant Thermus aqu
Synthesis deficient
5' Nuclease Cleava
Thermus aquatilis
Nucleotide sequenc
DNA sequence of a
Cleavease BN/thromb
Cleavease BN/thromb
Nucleotide sequenc

PT	Treating glomerulonephritis with antibody against complement C5 component - to inhibit complement induced cell lysis
XX	
P5	Example 11; Page 141-146; 181pp; English.
XX	
SQ	A DNA construct (AA708488) coding for a humanised CDR-grafted scFv, designated 5G1.1 scFv D012 (AA77616), was subcloned into CC CC bacterial expression vector pET Trc S05/N1 and the resulting CC plasmid (sequence given in AA708489) was transformed into CC Escherichia coli ME1. Humanised, recombinant scFv was obtl. CC which retained the ability of Mab 5G1.1 to block human complement CC C5a and C5b-9 generation and which was thus able to reduce CC glomerular inflammation and kidney dysfunction associated with CC glomerulonephritits.
XX	
SQ	Sequence 5248 BP: 1226 A; 1373 C; 1381 G; 1268 T; 0 other;
QY	Query Match 8.2%; Score 65.6; DB 16; Length 5248; Best Local Similarity 88.8%; Pred. NO. 8.6e-10; Matches 71; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Db	725 gagatgaacaaactcgtccgaattcgagctcgctgcagaacagtgcgagcaccgcagc 5033 gagatcgtgatccatcagatcgaattcgagctcgctgcagaacagtgcgagcaccgcagc 5092
Oy	785 accaccaccaccaccacctga 804 5093 accaccaccaccaccacctga 5112
RESULT 2	
AA741898	
ID	AA741898 standard; DNA: 5248 BP.
XX	
AC	AA741898;
XX	
DT	01-FEB-1997 (first entry)
XX	
DE	Vector pET Trc S05/N1.
XX	
KW	Vector: pET Trc S05.N1; myelin basic protein; MBP; K proteinlipid protein; PLP; multiple sclerosis; autoimmune disease; diagnosis; therapy; T-lymphocyte; T-cell; energy; apoptosis; ds; cyclic.
KM	
XX	
OS	Synthetic.
XX	
PN	WO9634622-A1.
XX	
PD	07-NOV-1996.
XX	
PE	22-APR-1996; 96WO-US05611.
XX	
PR	07-JUN-1995; 95US-0482114. 02-MAY-1995; 95US-0431644. 02-MAY-1995; 95US-0431648.
XX	
PA	(ALEX-) ALEXION PHARM INC. (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Lemardo MJ, Mats L, McFarland HF, Mueller EE, Mueller JP; Nye SH, Peltrey CM, Squinto SP, Wilkins JA;
XX	
DR	WPI: 1996-505898/50.
XX	
PT	New human myelin basic protein and proteolipid protein variant(s) - used in the assessment, diagnosis and treatment of multiple sclerosis
XX	
PS	Disclosure; Page 100-103; 156pp; English.
XX	
CC	Plasmid pET Trc S05/N1 (AA741898) is a bacterial expression vector.

```

CC      It is suitable for expression of human native and modified myelin
CC      basic protein and proteolipid protein (see also AA000399-400 and
CC      AA006101-08) in bacterial host cells, e.g. Escherichia coli.
XX
SQ      Sequence 5248 BP; 1226 A; 1373 C; 1381 G; 1268 T; 0 other;

Query Match          8.2%; Score 65.6; DB 17; Length 5248;
Best Local Similarity 88.8%; Pred. No. 8,6e-10;
Matches 71; Conservative 0; Mismatches 9; Indels 0; Gaps 0

Oy      725 gagatgcagaccgcgcgcgaattcgagctcgtcgacaaagcttcgagcgccactcgagc 784
       ||||| |||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      5033 gagatcgcgatcatcgatgatgaattcgagctcgcgcgacaaagcttcgagcgccactcgagc 5092
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      785 accacacacacacacacactga 804
       ||||| ||||| ||||| |||||
Db      5093 accacacacacacacacactga 5112
       ||||| ||||| ||||| |||||

RESULT 3
ID      AAX28088/c
XX      AAX28088 standard; DNA: 4411 BP.
AC      AAX28088;
XX
DT      10-JUN-1999 (first entry)
XX
DE      Plasmid PACYC.tac.
XX
KW      MuvV reverse transcriptase; dsCDNA production; cDNA library production;
XX      SS.
XX      Synthetic.
XX      OS
XX      US5891637-A.
XX      06-APR-1999.
XX      15-SEP-1997; 97US-0929967.
XX      15-OCR-1996; 96US-0732861.
XX      15-SEP-1997; 97US-0929967.
XX      (GETH ) GENENTECH INC.
XX      PA
XX      Ruppert SJ;
XX      WPI; 1999-253852/21.
XX      DR
XX      PT
XX      PT novel xenotropic murine leukemia virus reverse transcriptase -
XX      useful for constructing cDNA libraries
XX      Example 5; Fig 9; 70pp; English.
XX
CC      This sequence represents a plasmid used to test the method of
CC      the invention. The method is for the production of double
CC      stranded cDNA (dsCDNA) molecules from mRNA transcripts and a xenotropic
CC      murine leukemia virus reverse transcriptase (XM-MuV). The method, the
CC      host cells, reverse transcriptases and nucleic acids disclosed may be
CC      used to produce high quality, normalized, full-length, directionally
CC      cloned, cDNA libraries. Use of the method to construct cDNA libraries
CC      ensures that the reverse transcriptase is able to extend the first strand of
CC      the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
CC      all sequences are represented in the library, whereas previously, some
CC      sequences would be lost due to incomplete transcription (e.g. as a result
CC      of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
CC      H activity). Also, carrying the process out in cells means that enzymes
CC      and proteins present in the cell will repair any mistakes or nicks in the
CC      product DNA. Modifying the mRNA for insertion into a vector by adding the
CC      5' cap provides a convenient way of priming it for cDNA synthesis. Also,
CC      the method provides a fast and clean way of synthesizing cDNA

```

```

CC clones.
XX
SO Sequence 4411 BP; 1065 A; 1156 C; 1124 G; 1066 T; 0 other;

Query Match      .8.1%; Score 64.8; DB 20; Length:4411;
Best Local Similarity 97.1%; Pred.No.1.4e-09;
Matches 66; Conservative 0; Mismatches 2; Indels 0; Gaps 0

QY 737 ctgctcgaattcgaagctccgtcgcacaaagctgcgagcactcgagccaccacacc 796
   1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
Db 1719 CGGATCCGAATTCGAGCTCCGTCGACAGGCTTGCGGCGCACTCGAGCCACCAACCACC 1660
   1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

QY 797 accactga 804
   1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
Db 1659 ACCACTGA 1652

RESULT 4
AAx28085
ID AAx28085 standard; DNA; 5616 BP.
XX
AC AAx28085;
XX
DT 10-JUN-1999 (first entry)
XX
DE Plasmid PET21.tac.
XX
KW Mulv reverse transcriptase; dsCDNA production; cDNA library production;
KM ss.
XX
OS Synthetic.
XX
PN US5891637-A.
PD 06-APR-1999.
XX
PF 15-SEP-1997; 97US-0929967.
XX
PR 15-OCT-1996; 96US-0732861.
PR 15-SEP-1997; 97US-0929967.
XX
PA (GETH ) GENENTECH INC.
PI Ruppert SJ;
PI
DR WPI: 1999-253852/21.
XX
XX
XX Producing double stranded cDNA molecules from mRNA transcripts and a
XX novel xenotropic murine leukemia virus reverse transcriptase -
XX useful for constructing cDNA libraries
XX
XX Example 5; Fig 7; 70pp; English.
XX
XX This sequence represents a plasmid used to test the method of
XX the invention. The method is for the production of double
XX stranded cDNA (dsCDNA) molecules from mRNA transcripts and a xenotropic
XX murine leukaemia virus reverse transcriptase (XM-MULV). The method, the
XX host cells, reverse transcriptases and nucleic acids disclosed may be
XX used to produce high quality, normalized, full-length, directionally
XX cloned, cDNA libraries. Use of the method to construct cDNA libraries
XX ensures that the reverse transcriptase is able extend the first strand of
XX the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
XX all sequences are represented in the library, whereas previously, some
XX sequences would be lost due to incomplete transcription (e.g. as a result
XX of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
XX H activity). Also, carrying the process out in cells means that enzymes
XX and proteins present in the cell will repair any mistakes or nicks in the
XX product DNA. Modifying the mRNA for insertion into a vector by adding the
XX 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
XX the method provides a fast and clean way of synthesizing cDNA
XX clones.
XX

```

DR P-PSDB; AAM07193;
XX P-PSDB; AAM07194.
XX Vaccine for inducing mucosal response to Helicobacter - contg.
PT multimeric urease complex and pref. an antibiotic, anti-secretory
PT agent or bismuth salt
XX
XX
XX Disclosure; Page 68-70; 98pp; English.
PS
XX A cDNA clone (AAT44351), derived from PORV214, includes the urea and
CC urea sequences coding for the urease A (AAM07193) and urease B
CC (AAM07194) subunits of Helicobacter pylori clinical isolate CPM630.
CC To obtain PORV214, a genomic DNA library of CPM630 was screened
CC with anti-Helicobacter urease antibody. A 17 kb SalI fragment from
CC an isolated clone was subcloned into pUC18 to give pSCP1. PCR
CC primers (AAT44352-53) were used to amplify a 2.5 kb fragment from
CC pSCP1, which was inserted into pET24+ to give PORV214. The vector
CC was utilized in the prodn. of recombinant, enzymatically inactive,
CC multimeric urease in E. coli transformants for use in vaccines to
CC treat or prevent Helicobacter infection.
XX
XX Sequence 2735 BP; 875 A; 562 C; 635 G; 663 T; 0 other;

Query Match 7.6%; Score 61.4; DB 17; Length 2735;
Best Local Similarity 98.4%; Pred. No. 1.2e-08;
Matches 62; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 742 cgaattcgagctccgtcgacaaagcttgccgacgtcgagcaccacccaccacac 801
DB 2562 cgaattcgagctccgtcgacaaagcttgccgacgtcgagcaccacccaccac 2621
OY 802 tga 804
DB 2622 tga 2624

RESULT 6
AAX80369
ID AAX80369 standard; cDNA; 2007 BP.
XX
XX AAX80369;
AC
XX
DT 07-SEP-1999 (first entry)
XX
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:119.
DE
XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX WO9928482-A2.
PN
XX
XX 10-JUN-1999.
PD
XX
XX 24-NOV-1998; 98WO-US24528.
PF
XX
XX 28-JUL-1998; 98US-0094331.
PR
XX 28-NOV-1997; 97US-0067315.
PR
XX
XX (SCHE) SCHERING CORP.
PA
XX
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
PI
XX
XX WPI: 1999-385385/32.
DR
XX
XX New hepatitis C virus covalent complexes
PT
XX
XX Disclosure; Page 204-207; 211pp; English.
PS
XX

CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
XX Sequence 2007 BP; 413 A; 621 C; 563 G; 410 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctccgtcgacaaagcttgccgacgtcgagcaccacccaccacactg 803
DB 1942 gaattcgagctccgtcgacaaagcttgccgacgtcgagcaccacccaccacactg 2001
OY 804 a 804
DB 2002 a 2002

RESULT 7
AAX80370
ID AAX80370 standard; cDNA; 2007 BP.
XX
XX AAX80370;
AC
XX
DT 07-SEP-1999 (first entry)
XX
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:120.
DE
XX
XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX WO9928482-A2.
PN
XX
XX 10-JUN-1999.
PD
XX
XX 24-NOV-1998; 98WO-US24528.
PF
XX
XX 28-JUL-1998; 98US-0094331.
PR
XX 28-NOV-1997; 97US-0067315.
PR
XX
XX (SCHE) SCHERING CORP.
PA
XX
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
PI
XX
XX WPI: 1999-385385/32.
DR
XX
XX New hepatitis C virus covalent complexes
PT
XX
XX Disclosure; Page 207-210; 211pp; English.
PS
XX

CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

XX
SQ Sequence 2007 BP; 413 A; 622 C; 563 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgcgtgacaaagcttgcgcgcacactgcagacaccaccaccaccactg 803
|||||

DB 1942 gaattcgagctcgcgtgacaaagcttgcgcgcacactgcagacaccaccaccaccactg 2001

OY 804 a 804
DB 2002 a 2002

RESULT 8

AAx80366 standard; cDNA; 2007 BP.

XX
AC AAX80366;

XX
DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ. ID NO:116.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

PN MO9928482-A2.

XX
PD 10-JUN-1999.

XX
PE 24-NOV-1998; 98WO-US24528.

XX
PR 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX
PA (SCHE) SCHERING CORP.

XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX
DR WPI; 1999-385385/32.

XX
PT New hepatitis C virus covalent complexes

XX
PS Disclosure; Page 194-197; 211pp; English.

XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

XX
SQ Sequence 2007 BP; 415 A; 619 C; 564 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgcgtgacaaagcttgcgcgcacactgcagacaccaccaccaccactg 803
|||||

DB 1942 gaattcgagctcgcgtgacaaagcttgcgcgcacactgcagacaccaccaccaccactg 2001

OY 804 a 804
DB 2002 a 2002

RESULT 9

AAx80367 standard; cDNA; 2007 BP.

XX
AC AAX80367;

XX
DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ. ID NO:117.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

PN MO9928482-A2.

XX
PD 10-JUN-1999.

XX
PE 24-NOV-1998; 98WO-US24528.

XX
PR 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX
PA (SCHE) SCHERING CORP.

XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX
DR WPI; 1999-385385/32.

XX
PT New hepatitis C virus covalent complexes

XX
PS Disclosure; Page 197-200; 211pp; English.

XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

XX
SQ Sequence 2007 BP; 415 A; 619 C; 564 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgcgtgacaaagcttgcgcgcacactgcagacaccaccaccaccactg 803
|||||

DB 1942 gaattcgagctcgcgtgacaaagcttgcgcgcacactgcagacaccaccaccaccactg 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:48:29 ; Search time 1874.7 Seconds

(without alignments)
4054.032 Million cell updates/sec

Title: US-09-654-652A-5

Perfect score: 804
Sequence: 1 atggttagcgcaagattt.....accaccaccacacactga 804

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est12.*
3: gb_est13.*
4: gb_est14.*
5: gb_est15.*
6: gb_est16.*
7: gb_est17.*
8: gb_est18.*
9: gb_est19.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*

44: em_est10.*
45: em_est11.*
46: em_est12.*
47: em_est13.*
48: em_est14.*
49: em_est15.*
50: em_est16.*
51: em_est17.*
52: em_est18.*
53: em_est19.*
54: em_est20.*
55: em_est21.*
56: em_est22.*
57: em_est23.*
58: em_est24.*
59: em_est25.*
60: em_est26.*
61: em_est27.*
62: em_est28.*
63: em_est29.*
64: em_est30.*
65: em_est31.*
66: em_est32.*
67: em_est33.*
68: em_est34.*
69: em_est35.*
70: em_est36.*
71: em_est37.*
72: em_est38.*
73: em_est39.*
74: em_est40.*
75: em_est41.*
76: em_est42.*
77: em_est43.*
78: em_est44.*
79: em_est45.*
80: em_est46.*
81: em_est47.*
82: em_est48.*
83: em_est49.*
84: em_est50.*
85: em_est51.*
86: em_est52.*
87: em_est53.*
88: em_est54.*
89: em_est55.*
90: em_est56.*
91: em_est57.*
92: em_est58.*
93: em_est59.*
94: em_est60.*
95: em_est61.*
96: em_est62.*
97: em_est63.*
98: em_est64.*
99: em_est65.*
100: em_est66.*
101: em_est67.*
102: em_est68.*
103: em_est69.*
104: em_est70.*
105: em_est71.*
106: em_est72.*
107: em_est73.*
108: em_est74.*
109: em_est75.*
110: em_est76.*
111: em_est77.*
112: em_est78.*
113: em_est79.*
114: em_est80.*
115: em_est81.*
116: em_est82.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47	5.8	599	155	BG605030	WHE2326-D
2	45.4	5.6	529	136	BE494282	WHE1253-H
3	41	5.1	418	136	BE495688	WHE1279-B
4	37.8	4.7	796	15	A1068991	mgae0005a
5	37.8	4.7	849	15	A1069170	mgae0005d
6	37.2	4.6	561	13	BG103002	RH122-35-
7	37	4.6	552	19	A1397807	NCM9C4T3
8	36.8	4.6	747	170	BF661648	BF861648 963025E08
9	36.4	4.5	408	158	H92520	ys85a11.r1
10	36.4	4.5	604	166	BE359676	DC1.56.H0
11	36.4	4.5	608	154	BG488141	RH122.60-
12	36.2	4.5	635	167	BE405426	WHE1216-C
13	35.8	4.5	587	145	BF202533	WHE1776-G
14	35.8	4.5	584	133	AA898069	NCM3G7V7
15	35.8	4.5	587	133	AA897960	NCM1H9T7
16	35.4	4.4	399	237	AZ046873	nbe00088H
17	35.4	4.4	640	141	BE888160	601511671
18	35.4	4.4	669	164	BE208885	GF-EV-P3C
19	35.4	4.4	944	169	BF782886	602107758
20	35.2	4.4	360	167	BE428429	MTD006.H1
21	35.2	4.4	423	167	BE398687	WHE0024.C
22	35.2	4.4	536	149	BE473353	WHE0923.E
23	35.2	4.4	552	167	BE419465	WMS012.D8
24	35.2	4.4	776	151	BF627470	HVSMED000
25	35	4.4	340	116	AW486705	76371.MAR
26	35	4.4	482	166	BE347923	SP08H02.Y
27	35	4.4	495	166	BE059962	sn39C06.Y
28	35	4.4	649	230	AO575231	npxb0087A
29	35	4.4	762	170	BF665955	963063D08
30	35	4.4	781	141	BE884266	601505733
31	34.8	4.3	826	146	BF283101	HV-CEA000
32	34.6	4.3	876	150	BF526437	602070921
33	34.4	4.3	443	104	A1940919	SB1A11.Y
34	34.4	4.3	474	33	AV668531	NCM6G5T7
35	34.2	4.3	376	144	BF072772	NCM5G11T
36	34.2	4.3	455	175	BE278686	ade05np.f
37	34.2	4.3	501	175	BE279070	b2f05np.f
38	34.2	4.3	521	175	BE279385	ab970np.f
39	34.2	4.3	532	175	BE279532	b3911np.f
40	34.2	4.3	546	30	AV386476	AV386476
41	34.2	4.3	548	13	AA898699	NCM6G5T7
42	34.2	4.3	550	13	AA898546	NCM5G12T7
43	34.2	4.3	552	175	BE279542	b3h07np.f
44	34.2	4.3	558	13	AA898246	NCM4A9T7
45	34.2	4.3	560	175	BE280207	c1f11np.f

ALIGNMENTS

RESULT 1	BG605030	599 bp	mRNA	EST	16-APR-2001
LOCUS	WHE2326_D09.G1875	wheat pre-anthesis spike cDNA library	Triticum aestivum		
DEFINITION	BG605030	16-APR-2001			
ACCESSION	BG605030				
VERSION	BG605030.1	GI:13635033			
KEYWORDS	EST				
SOURCE	bread wheat.				
ORGANISM	Triticum aestivum				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae				
AUTHORS	Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.				

TITLE The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library

JOURNAL

Unpublished (2000)

COMMENT

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@wv.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: Stragene SK primer.

FEATURES

SOURCE

1..599

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db_xref="taxon:4565"

/clone="WHE2326_D09.G18"

/clone_lib="Wheat pre-anthesis spike cDNA library"

/tissue_type="Spike before anthesis"

/dev_stage="Adult plant"

/lab_host="E. coli SOLR"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phagescript phagemids in the T7 Close Lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT

141 a

182 c

156 g

120 t

ORIGIN

Query Match 5.8%; Score 47; DB 155; Length 599; Best Local Similarity 51.2%; Pred. No. 0.0018; Matches 110; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY	583	ggtccgcgtgaggaaggtgactgacattgacgtaacggtgactcaacgac	642
DB	83	GGCCCTCACCATTGTACACGAGGAGCGGAGCGGAGCGGCGCCGCCAAC	142
QY	643	aagaacatctactcagagatgcatgttgcctccgacctcgaagaaggtcagaa	702
DB	143	GCCACAGCCTTTCAGCGCGGCGGCTTCATGCTCGACACCAACACCCGGAG	202
QY	703	agctcaacgacaggttcgagagatgacgaacctgctcgaattcgagctcgac	762
DB	203	GGCAGCGGCGCATCACCAAGACTCGCGGCGGAGAGCGCTCTCCCGTCAAG	262
QY	763	aagcttgagccgacactgagacacacacacga	797
DB	263	CAGATTATGAGCGCTCCAGCACACGACGACGACAA	297

RESULT 2

LOCUS	BE494282	529 bp	mRNA	EST	02-AUG-2000
DEFINITION	WHE1253_H10.P1925	Secale cereale anther cDNA library	Secale cereale		
ACCESSION	BE494282				
VERSION	BE494282.1	GI:9660875			
KEYWORDS	EST				
SOURCE	Secale cereale				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae				
REFERENCE	1 (bases 1 to 529)				
AUTHORS	Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton				

TITLE
R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R.,
Miller, R., Rausch, C.J., Ross, K., Seaton, C.L. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Another CDNA library from rye

JOURNAL
Unpublished (2000)

COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stralagene SK primer.
Location/Qualifiers

FEATURES
source

1.529
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE1253_H10_P19"
/clone_lib="Secale cereale anther cDNA library"
/tissue_type="Anther"
/dev_stage="Adult plant before anthesis"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
greenhouse. Anthers were harvested and pooled from early
meiosis to late meiosis. The tissue, total RNA, and
poly(A) RNA were prepared (Butler, Ross and Gustafson) at
University of Missouri, Columbia. A cDNA library was
made, and the cDNA clones were in vivo excised to give
phagescript phagemids in the TJ Close lab (Choi, Close,
Fenton) at the University of California, Riverside.
Plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson lab (all other authors)."

BASE COUNT
127 a 153 c 149 g 100 t

ORIGIN

Query Match
Best Local Similarity 50.7%; Score 45.4; DB 136; Length 529;
Pred. No. 0.0053;
Matches 109; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 583 ggtccgcgtgggcaagggtgacatgtgacgtgacgtgtgacgtccaccgac 642
|||||
DB 61 GGCTCACTGATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
QY 643 aagaacatctactcagagatgcatgtgatctcgcctcaccgcaaggtcagaa 702
|||||
DB 121 GCCACAGCCTCTTCAAGCGGGGGGTTTCATGCCCTCGACACACACACCCCGGAG 180
QY 703 agctcaagcgaggttccgagatgacgaacctgtcgaattcgagctcgtagc 762
|||||
DB 181 GGCAGCGGGCGGCACTCAGCAAGACTCGGGCGGCGGAGACGCTGCTCCCGCTCAAG 240
QY 763 aagcttgcggcgaactcggagaccaccaccacca 797
|||||
DB 241 CAGATTATGACCGCTCCAGACCAAGAGAGACAA 275

RESULT
3
BE495688 418 bp mRNA EST 02-AUG-2000
LOCUS BE495688
DEFINITION WHE1279_B01_C01ZS Secale cereale anther cDNA library Secale cereale
CDNA clone WHE1279_B01_C01, mRNA sequence.
ACCESSION BE495688
VERSION BE495688.1 GI:9662281
KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

REFERENCE
1 (bases 1 to 418)
Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton
R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R.,
Miller, R., Rausch, C.J., Ross, K., Seaton, C.L. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Another CDNA library from rye

JOURNAL
Unpublished (2000)

COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stralagene SK primer.
Location/Qualifiers

FEATURES
source

1.418
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE1279_B01_C01"
/clone_lib="Secale cereale anther cDNA library"
/tissue_type="Anther"
/dev_stage="Adult plant before anthesis"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
greenhouse. Anthers were harvested and pooled from early
meiosis to late meiosis. The tissue, total RNA, and
poly(A) RNA were prepared (Butler, Ross and Gustafson) at
University of Missouri, Columbia. A cDNA library was
made, and the cDNA clones were in vivo excised to give
phagescript phagemids in the TJ Close lab (Choi, Close,
Fenton) at the University of California, Riverside.
Plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson lab (all other authors)."

BASE COUNT
77 a 180 c 103 g 58 t

ORIGIN

Query Match
Best Local Similarity 50.8%; Score 41; DB 136; Length 418;
Pred. No. 0.11;
Matches 98; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 583 ggtccgcgtgggcaagggtgacatgtgacgtgacgtgtgacgtccaccgac 642
|||||
DB 90 GGCTCTCTCTTGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149
QY 643 aagaacatctactcagagatgcatgtgatctcgcctcaccgcaaggtcagaa 702
|||||
DB 150 AAAAGATGACCTCACCAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 209
QY 703 agctcaagcgaggttccgagatgacgaacctgtcgaattcgagctcgtagc 762
|||||
DB 210 GCCGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 269
QY 763 aagcttgcggcg 775
|||||
DB 270 CACCTCGCCATCG 282

RESULT
4
A1068991 796 bp mRNA EST 09-DEC-1999
LOCUS A1068991
DEFINITION mgae0005ab09f Magnaporthe grisea Appressorium Stage cDNA library
Magnaporthe grisea cDNA clone mgae0005ab09f 5', mRNA sequence.
ACCESSION A1068991
VERSION A1068991.1 GI:3391966
KEYWORDS EST.
SOURCE Magnaporthe grisea.

ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 796)
AUTHORS Choi, W., Fang, E., Sasinowski, M., Wing, R. and Dean, R.A.
TITLE Expressed sequence characterization during appressorium formation
in rice blast fungus, Magnaporthe grisea
JOURNAL Unpublished (1998)
COMMENT Contact: Dean, R.A.
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdeane@clemson.edu
Seq primer: T3 primer (AATTAACCCCTCACTAAGG)
High quality sequence stop: 348.
Location/Qualifiers

FEATURES
source
1. .796
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgae0005ab09f"
/clone_lib="Magnaporthe grisea Appressorium Stage cDNA
Library"
/dev_stage="Germinated conidia on appressorium-inductive
surface"
/note="Vector: pBluescriptII SK(+). Vector: Site.1: EcoRI;
Site.2: XhoI. The appressorium formation-specific cDNA
library was constructed from conidia germinated for 5-8
hr on an inductive surface. The library has an average
insert size of 1.5 kbp."
BASE COUNT 175 a 268 c 198 g 146 t 9 others
ORIGIN

Query Match 4.7%; Score 37.8; DB 15; Length 796;
Best Local Similarity 49.7%; Pred. No. 1.3;
Matches 96; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

185 tcggcaagaatccggcaggttccagtcacacatcattaccggttaagccggcgacaaa 244
|||||
166 tcggaaacgacacagatcggtctccagacacacttctccaaaggtaatgacacgctct 225
|||||
245 agactagcgaagaacacacatgctgttagcccgccgcgcgataagcttccacacctag 304
|||||
226 atggccctggccagttccacagactcctctgccaacgacgacatgacacactcctctgac 285
|||||
305 gtctcgaatgactccgaattacgctcgcctgagactgttgcagtcaggaagtcgcgaaga 364
|||||
286 ctcttgactgagaccagacgacgctcagtgatgttgcacggcgaagtcgtccgacacc 345
|||||
365 cggagagtgacca 377
|||||
346 tgaagcgcgcgca 358
|||||

RESULT 5
LOCUS A1069170 849 bp mRNA EST 09-DEC-1999
DEFINITION mgae0005dC07f Magnaporthe grisea Appressorium Stage cDNA Library
Magnaortha grisea cDNA clone mgae0005dC07f 5', mRNA sequence.
ACCESSION A1069170
VERSION A1069170.1 GI:3392145
KEYWORDS EST.
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 849)
AUTHORS Choi, W., Fang, E., Sasinowski, M., Wing, R. and Dean, R.A.
TITLE Expressed sequence characterization during appressorium formation
in rice blast fungus, Magnaporthe grisea

JOURNAL Unpublished (1998)
COMMENT Contact: Dean, R.A.
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdeane@clemson.edu
Seq primer: T3 primer (AATTAACCCCTCACTAAGG)
High quality sequence stop: 324.
Location/Qualifiers

FEATURES
source
1. .849
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgae0005dC07f"
/clone_lib="Magnaporthe grisea Appressorium Stage cDNA
Library"
/dev_stage="Germinated conidia on appressorium-inductive
surface"
/note="Vector: pBluescriptII SK(+). Vector: Site.1: EcoRI;
Site.2: XhoI. The appressorium formation-specific cDNA
library was constructed from conidia germinated for 5-8
hr on an inductive surface. The library has an average
insert size of 1.5 kbp."
BASE COUNT 203 a 290 c 198 g 152 t 6 others
ORIGIN

Query Match 4.7%; Score 37.8; DB 15; Length 849;
Best Local Similarity 50.8%; Pred. No. 1.4;
Matches 90; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

185 tcggcaagaatccggcaggttccagtcacacatcattaccggttaagccggcgacaaa 244
|||||
196 tcggaaacgacacagatcggtctccagacacacttctccaaaggtaatgacacgctct 255
|||||
245 agactagcgaagaacacacatgctgttagcccgccgcgcgataagcttccacacctag 304
|||||
256 atggccctggccagttccacagactcctctgccaacgacgacatgacacactcctctgac 315
|||||
305 gtctcgaatgactccgaattacgctcgcctgagactgttgcagtcaggaagtcgcga 361
|||||
316 ctcttgactgagaccagacgacgctcagtgatgttgcacggcgaagtcgtccgca 372
|||||

RESULT 6
LOCUS BG103002 561 bp mRNA EST 30-JAN-2001
DEFINITION RH122.35_H06_b1_A003 Rhizome2 (RH122) Sorghum prolingum cDNA, mRNA
sequence.
ACCESSION BG103002
VERSION BG103002.1 GI:12617835
KEYWORDS EST.
SOURCE Sorghum prolingum.
ORGANISM Sorghum prolingum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 561)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt
L.H.
TITLE An EST database from Sorghum: Sorghum prolingum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 550

FEATURES

source

POLYA-No.

Location/Qualifiers

1. .561
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II; clones to be sequenced were prepared by mass excision."

BASE COUNT 101 a 180 c 196 g 84 t

ORIGIN

Query Match 4.6%; Score 37.2; DB 173; Length 561;
Best Local Similarity 49.0%; Pred. No. 1.8;

Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 161 gggtagaagtgagatattgaagtcctcgcaagaatccggcagttccagttccacaatca 220

Db 8 GGCACAGATCGACTTCGAGTCTCCGCAACGTCGCCGCGAGCCCTACACGCTGCACA 67

Qy 221 ttacggtgaagcggcgcaacaaagactagcgaaagacacatgtctgttagcccgccg 280

Db 68 CCACAGCTTTCACCGCGCGCGACGCGCACGCGAGCGATTCGCCCTCTGTTGACACC 127

Qy 281 ccgataagcttcacacactacagtcgtcgaatgagctccgaattacgttcgtgagctg 340

Db 128 CCACCAAGGACTTCCACACTTACTCCGTGTGTGGAACCCGACGACGATGATCTTGCCG 187

Qy 341 ttgacggtcagaagatccgcaa 362

Db 188 TGGACGCGACCCCGATCCGCGA 209

RESULT 7

LOCUS

AI397807 552 bp mRNA EST 10-JAN-2001

DEFINITION NC_00473 Mycelial Neurospora crassa cDNA clone NM9C4 5', mRNA

ACCESSION AI397807

VERSION AI397807.1 GI:4240892

KEYWORDS EST.

SOURCE Neurospora crassa.

ORGANISM Neurospora crassa

REFERENCE 1 (bases 1 to 552)

AUTHORS Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,

Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E.,

Cushling, T., Errett, A., Flenahy, M., Gorman, M., Judson, K., Miller, R.,

Ortega, J., Pavlova, I., Perea, J., Todisco, S., Trujillo, R.,

Valentine, J., Wells, A., Werner, Washburne, M., Yazzie, S. and Nativg

, D.O. Expressed sequences from conidial, mycelial, and sexual stages of

Neurospora crassa

JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)

MEDLINE 97435549

COMMENT Contact: Nativg, D.O./Nelson, M.A.

Department of Biology

University of New Mexico

Casletter Hall, Albuquerque, NM 87131, USA

Tel: 505 277 3411

Fax: 505 277 0304

Email: npebbiology.unm.edu.

Location/Qualifiers

1. .552

/organism="Neurospora crassa"

/strain="74-OR23-IV A (FGSC 2489)"

/db_xref="taxon:5141"

/clone="NM9C4"

/clone_lib="Mycelial"

/sex="Mating type A"

FEATURES

source

POLYA-No.

Location/Qualifiers

1. .561
/tissue_type="Mycelium"
/dev_stage="Mycelium"
/lab_host="E. coli"
/note="Vector: pBluescript SK (-); Site_1: EcoRI; Site_2: XhoI; 2% sucrose for 24 hours. cDNA directionally cloned into pBluescript SK(-) using the Uni-ZAP XR vector system (Stratagene, La Jolla, CA)."

BASE COUNT 127 a 137 c 172 g 116 t

ORIGIN

Query Match 4.6%; Score 37; DB 19; Length 552;
Best Local Similarity 46.9%; Pred. No. 2.1;

Matches 115; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

Qy 455 tcgataatcaaaagcttcgcttcagttccagttcaactcagttcagttataatata 514

Db 293 TCGAAGACTACGTGATCCGATCGACAGCGTCACACTCCGCGAGACTACGCGATCG 352

Qy 515 ccgagggcagagcgagagcgagcagacttaacgttactgagaccgacaatttga 574

Db 353 GGGTCGACAAAGCGTGTGCGCTCGAGAGACTATGTATGAGGTGCGACAGCGTCA 412

Qy 575 cgttatgtctccgctggtggcaaggtgactgacattgacgttgaacgtgtgcac 634

Db 413 GTGTTGAGGACTACGCGCATCGGGGTGATTAAGGTGTTCTGCGAGGACTACG 472

Qy 635 taaccgacagaacatctactcagagatgcatgtgtactcgtccaccacccgaag 694

Db 473 CGATCGACAAAGCGCGACACTCCGTCGAAGACTATGTATGAAGTCGACAGTCT 532

RESULT 8

LOCUS

BF861648 747 bp mRNA EST 19-JAN-2001

DEFINITION 963025E08.Y1 C. reinhardtii CC-1690, stress condition I, normalized

ACCESSION BF861648

VERSION BF861648.1 GI:12251785

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii

REFERENCE 1 (bases 1 to 747)

AUTHORS Grossman, A., Davies, J., Federapfel, N., Harris, E., Hauser, C.,

Lefebvre, P., McDermott, J.P., Shrago, J., Silflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; project phase 3

JOURNAL unpublished (2000)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1. .747

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, stress condition I,

normalized, Lambda zap II"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:

XhoI; This library, constructed by John Davies and Jeffrey

McDermott, combines cDNAs from CC-1690 cells grown to

mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,

Fri Aug 24 10:05:31 2001

us-09-654-652a-5.rst

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 15:04:24 ; Search time 2906.88 Seconds
(without alignments)
4278.151 Million cell updates/sec

Title: US-09-654-652A-5
Perfect score: 804
Sequence: 1 atggttagcgcaaggattt.....accaccaccaccaccactga 804

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_p11:*
13: gb_p12:*
14: gb_p13:*
15: gb_p14:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_p1:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rol2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	741	92.2	1426	3	FTBLGUC
2	64	8.0	4411	9	AR069366
3	64	8.0	5443	9	AR069362
4	64	8.0	5616	9	AR069364
5	61	7.6	2735	9	AR054309
6	55	6.8	6363	9	AR069367
7	45	5.6	98	9	AX026763
8	43	5.3	117	56	ASTOOL10

Db 937 AGAGATGACGAACTGCTCG 957

RESULT 2
LOCUS AR069366/c 4411 bp DNA
DEFINITION Sequence 5 from patent US 5891637.
ACCESSION AR069366
VERSION AR069366.1 GI:7220254
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4411)
AUTHORS Ruppert,S.J.W.
TITLE Construction of full length cDNA libraries
JOURNAL Patent: US 5891637-A 5 06-APR-1999;
FEATURES
Source 1..4411
Location/Qualifiers
BASE COUNT 1065 a 1156 c 1124 g 1066 t
ORIGIN

Query Match 8.0%; Score 64; DB 9; Length 4411;
Best Local Similarity 100.0%; Pred. No. 3.1e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcagaaagcttgcggccgacactgcagagaccacacacacca 800
|||||
Db 1715 TCGGAATTGCGAGCTCCGTCGACAGCTTGCGCGCGCAGCTGAGACACACACACACCA 1656

QY 801 ctga 804
|||||
Db 1655 CTGA 1652

RESULT 3
LOCUS AR069362 5443 bp DNA
DEFINITION Sequence 1 from patent US 5891637.
ACCESSION AR069362
VERSION AR069362.1 GI:7220250
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5443)
AUTHORS Ruppert,S.J.W.
TITLE Construction of full length cDNA libraries
JOURNAL Patent: US 5891637-A 1 06-APR-1999;
FEATURES
Source 1..5443
Location/Qualifiers
BASE COUNT 1268 a 1441 c 1446 g 1288 t
ORIGIN

Query Match 8.0%; Score 64; DB 9; Length 5443;
Best Local Similarity 100.0%; Pred. No. 3e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcagaaagcttgcggccgacactgcagagaccacacacacca 800
|||||
Db 5244 TCGGAATTGCGAGCTCCGTCGACAGCTTGCGCGCGCAGCTGAGACACACACACACCA 5303

QY 801 ctga 804
|||||
Db 5304 CTGA 5307

RESULT 4

AR069364
LOCUS AR069364 5616 bp DNA
DEFINITION Sequence 3 from patent US 5891637.
ACCESSION AR069364
VERSION AR069364.1 GI:7220252
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5616)
AUTHORS Ruppert,S.J.W.
TITLE Construction of full length cDNA libraries
JOURNAL Patent: US 5891637-A 3 06-APR-1999;
FEATURES
Source 1..5616
Location/Qualifiers
BASE COUNT 1306 a 1478 c 1492 g 1340 t
ORIGIN

Query Match 8.0%; Score 64; DB 9; Length 5616;
Best Local Similarity 100.0%; Pred. No. 3e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcagaaagcttgcggccgacactgcagagaccacacacacca 800
|||||
Db 5417 TCGGAATTGCGAGCTCCGTCGACAGCTTGCGCGCGCAGCTGAGACACACACACACCA 5476

QY 801 ctga 804
|||||
Db 5477 CTGA 5480

RESULT 5
LOCUS AR054309 2735 bp DNA
DEFINITION Sequence 1 from patent US 5837240.
ACCESSION AR054309
VERSION AR054309.1 GI:5979886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2735)
AUTHORS Lee,C.K., Monath,T.P., Ackerman,S.K., Thomas,W.D., Soman,G., Kleenhouus,H., Weitzin,R.A., Pappo,J., Ermak,T., Guirakhoo,F., Bhagat,H. and Sussman,I.
TITLE Multimeric, recombinant urease vaccine
JOURNAL Patent: US 5837240-A 1 17-NOV-1998;
FEATURES
Source 1..2735
Location/Qualifiers
BASE COUNT 875 a 562 c 635 g 663 t
ORIGIN

Query Match 7.6%; Score 61; DB 9; Length 2735;
Best Local Similarity 100.0%; Pred. No. 1.3e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 gaattcgagctcgctgcagaaagcttgcggccgacactgcagagaccacacacacactg 803
|||||
Db 2564 GAATTGAGCTCCGTCGACAGCTTGCGCGCGCAGCTGAGACACACACACACACTG 2623

QY 804 a 804
|
Db 2624 A 2624

RESULT 6
LOCUS AR069367 6363 bp DNA
DEFINITION Sequence 6 from patent US 5891637.
PAT 18-FEB-2000

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 caagcttgccgcgacactgcagcaccacacacacactga 804
|||||
Db 366 CAAGCTTGGCGCGCAGCTCGAGCAGCACCACACACCACTGA 408

RESULT 10
AF097413 5549 bp DNA circular SYN 20-APR-1999
LOCUS
DEFINITION Expression vector PHIS-parallel1, complete sequence.
ACCESSION AF097413
VERSION AF097413.1 GI:4588089
KEYWORDS
SOURCE Expression vector PHIS-parallel1.
ORGANISM Expression vector PHIS-parallel1.
REFERENCE 1 (bases 1 to 5549)
AUTHORS Sheffield,P., Garrard,S. and Derewenda,Z.
TITLE Overcoming expression and purification problems of RhodPI using a family of 'parallel' expression vectors
JOURNAL Protein Expr. Purif. 15 (1), 34-39 (1999)
MEDLINE 99150479
REFERENCE 2 (bases 1 to 5549)
AUTHORS Sheffield,P.J., Garrard,S.M. and Derewenda,Z.S.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Molecular Physiology and Biological Physics, University of Virginia, 4215 Jordan Hall, 1300 Jefferson Park Avenue, Charlottesville, VA 22908, USA
FEATURES
source Location/Qualifiers
1..5549
/organism="Expression vector PHIS-parallel1"
/db_xref="taxon:91376"

BASE COUNT 1297 a 1477 c 1463 g 1312 t

ORIGIN

Query Match 5.3%; Score 43; DB 56; Length 5549;
Best Local Similarity 100.0%; Pred. No. 2.8e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 caagcttgccgcgacactgcagcaccacacacacactga 804
|||||
Db 5371 CAAGCTTGGCGCGCAGCTCGAGCAGCACCACCACTGA 5413

RESULT 11
ASTCELL90 306 bp DNA SYN 08-NOV-1996
LOCUS
DEFINITION Artificial sequence DNA for T-cell receptor, toolbox 90.
ACCESSION Y092275
VERSION Y092275.1 GI:1666290
KEYWORDS T-cell receptor; toolbox 90.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 306)
AUTHORS Schluesener,H.J.
TITLE Unpublished
JOURNAL 2 (bases 1 to 306)
REFERENCE Schluesener,H.J.
AUTHORS Direct Submission
JOURNAL Submitted (21-OCT-1996) H.J. Schluesener, Inst.f. Hirnforschung, Calwer Str. 3, D-77076 Tuebingen, FRG
FEATURES
source Location/Qualifiers
1..306
/organism="synthetic construct"
/db_xref="taxon:32630"

CDS
1..306
/note="toolbox 90"
/codon_start=1
/transl_table=1
/product="T cell receptor peptide"

/protein_id="CAA70470.1"
/db_xref="GI:1666291"
/translation="MASMTGQOMGRGSEFAAAGSTLQOLFMSFVSGSEVELDM
GHGRLHYVSGSSPENGVDMDGHGRLHYSDVNSTEKGLEKLAALAEHHHHH"

BASE COUNT 65 a 103 c 86 g 52 t

ORIGIN

Query Match 5.2%; Score 42; DB 56; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 aagcttgccgcgacactgcagcaccacacacacactga 804
|||||
Db 265 AAGCTTGGCGCGCAGCTCGAGCAGCACCACCACTGA 306

RESULT 12
AX025780 357 bp DNA PAT 16-SEP-2000
LOCUS
DEFINITION Sequence 137 from Patent WO0028039.
ACCESSION AX025780
VERSION AX025780.1 GI:10187390
KEYWORDS
SOURCE SEN virus.
ORGANISM SEN virus
REFERENCE 1 (bases 1 to 357)
AUTHORS Viruses: ssDNA viruses; unclassified ssDNA viruses.
TITLE Flordalisi,G., Bonelli,M., Olivero,P., Primi,D., Vaglini,L.,
JOURNAL Mattioli,S., Bonelli,F., Dal,C.A., Mantero,G.L. and Sottini,A.
Identification of senv genotypes
Patent: WO 0028039-A 137 18-MAY-2000;
FIORDALISI GIANFRANCO (IT); BONELLI MARCO (IT); OLIVERO PAOLO
(IT); PRIMI DANIELE (IT); VAGLINI LAURA (IT); DIASORIN S R L
(IT); MATTIOLI SONIA (IT); BONELLI FABRIZIO (IT); DAL CORSO
ANDREA (IT); MANTERO GIOVANNI LORENZO (IT); SOTTINI ALESSANDRA
(IT)

FEATURES
source Location/Qualifiers
1..357
/organism="SEN virus"
/db_xref="taxon:136966"

BASE COUNT 117 a 109 c 76 g 55 t

ORIGIN

Query Match 5.2%; Score 42; DB 9; Length 357;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 aagcttgccgcgacactgcagcaccacacacacactga 804
|||||
Db 316 AAGCTTGGCGCGCAGCTCGAGCAGCACCACCACTGA 357

RESULT 13
AX025776 586 bp DNA PAT 16-SEP-2000
LOCUS
DEFINITION Sequence 133 from Patent WO0028039.
ACCESSION AX025776
VERSION AX025776.1 GI:10187387
KEYWORDS
SOURCE SEN virus.
ORGANISM SEN virus
REFERENCE 1 (bases 1 to 586)
AUTHORS Viruses: ssDNA viruses; unclassified ssDNA viruses.
TITLE Flordalisi,G., Bonelli,M., Olivero,P., Primi,D., Vaglini,L.,
JOURNAL Mattioli,S., Bonelli,F., Dal,C.A., Mantero,G.L. and Sottini,A.
Identification of senv genotypes
Patent: WO 0028039-A 133 18-MAY-2000;
FIORDALISI GIANFRANCO (IT); BONELLI MARCO (IT); OLIVERO PAOLO
(IT); PRIMI DANIELE (IT); VAGLINI LAURA (IT); DIASORIN S R L
(IT); MATTIOLI SONIA (IT); BONELLI FABRIZIO (IT); DAL CORSO
ANDREA (IT); MANTERO GIOVANNI LORENZO (IT); SOTTINI ALESSANDRA
(IT)

FEATURES	(IT)
source	Location/Qualifiers
	1. .586
	/organism="SEN virus"
	/db_xref="taxon:136966"
BASE COUNT	155 a 172 c 139 g 119 t
ORIGIN	1 others

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 15:01:15 ; Search time 157.23 Seconds
(without alignments)
3210.792 Million cell updates/sec

Title: US-09-654-652A-5

Perfect score: 804
Sequence: 1 atggttagcgcaagaattt.....accacacacacacactga 804

Scoring table:
OLIGO_NUC
Gapox 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

N_Geneseq_0601:*

1: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*

22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	64	8.0	4411	20	AAx28088
2	64	8.0	5616	20	AAx28085
3	61	7.6	2007	20	AAx80369
4	61	7.6	2007	20	AAx80370
5	61	7.6	2007	20	AAx80366
6	61	7.6	2007	20	AAx80367
7	61	7.6	2007	20	AAx80368
8	61	7.6	2735	17	AAx44351
9	61	7.6	5248	16	AAx08489
10	61	7.6	5248	17	AAx41898
11	55	6.8	6363	20	AAx28091

12	50	6.2	59	22	AAx74635
13	42	5.2	357	21	AAx61167
14	42	5.2	586	21	AAx61166
15	42	5.2	1186	20	AAx29726
16	42	5.2	1247	20	AAx29725
17	42	5.2	6619	20	AAx63789
18	41	5.1	423	20	AAx21936
19	41	5.1	423	20	AAx219184
20	39	4.9	550	21	AAx71987
21	37	4.6	607	21	AAx64807
22	36	4.5	241	21	AAx94064
23	36	4.5	1815	17	AAx39341
24	36	4.5	1815	19	AAx15138
25	34	4.2	948	16	AAx080748
26	34	4.2	948	17	AAx27687
27	34	4.2	948	18	AAx70349
28	34	4.2	948	18	AAx76649
29	34	4.2	948	19	AAx65799
30	34	4.2	948	19	AAx53873
31	34	4.2	948	20	AAx63425
32	34	4.2	963	18	AAx76626
33	34	4.2	963	19	AAx65831
34	34	4.2	963	19	AAx53941
35	29	3.6	783	16	AAx08490
36	28	3.5	738	22	AAx25015
37	28	3.5	981	21	AAx51210
38	28	3.5	6353	22	AAx68848
39	28	3.5	7375	21	AAx09770
40	28	3.5	7375	21	AAx98759
41	27	3.4	294	21	AAx51209
42	27	3.4	294	21	AAx51211
43	27	3.4	980	17	AAx42665
44	27	3.4	3572	20	AAx219454
45	27	3.4	3572	20	AAx219242

ALIGNMENTS

RESULT 1	
AAx28088/c	
ID	AAx28088 standard; DNA; 4411 BP.
AC	AAx28088;
XX	
XX	10-JUN-1999 (first entry)
XX	
DE	Plasmid PACYC.tac.
XX	
KW	Mult reverse transcriptase; cDNA production; cDNA library production; ss.
XX	
OS	Synthetic.
XX	
PN	US5891637-A.
XX	
PD	06-APR-1999.
XX	
PF	15-SEP-1997; 97US-0929967.
XX	
PR	15-OCT-1996; 96US-0732861.
PR	15-SEP-1997; 97US-0929967.
XX	
PA	(GETH) GENENTECH INC.
XX	
PI	Ruppert SJ;
XX	
DR	WPI; 1999-253852/21.
XX	
PT	Producing double stranded cDNA molecules from mRNA transcripts and a novel xenotropic murine leukemia virus reverse transcriptase - useful for constructing cDNA libraries
XX	

PS Example 5; Fig 9; 70pp; English.
XX
CC This sequence represents a plasmid used to test the method of
CC the invention. The method is for the production of double
CC stranded cDNA (dsCDNA) molecules from mRNA transcripts and a xenotropic
CC murine leukemia virus reverse transcriptase (xm-MuLV). The method, the
CC host cells, reverse transcriptases and nucleic acids disclosed may be
CC used to produce high quality, normalized, full-length, directionally
CC cloned, cDNA libraries. Use of the method to construct cDNA libraries
CC ensures that the reverse transcriptase is able to extend the first strand of
CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
CC all sequences are represented in the library, whereas previously, some
CC sequences would be lost due to incomplete transcription (e.g. as a result
CC of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
CC H activity). Also, carrying the process out in cells means that enzymes
CC and proteins present in the cell will repair any mistakes or nicks in the
CC product DNA. Modifying the mRNA for insertion into a vector by adding the
CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
CC the method provides a fast and clean way of synthesizing cDNA
CC clones.
XX
SQ Sequence 4411 BP; 1065 A; 1156 C; 1124 G; 1066 T; 0 other;

Query Match 8.0%; Score 64; DB 20; Length 4411;
Best Local Similarity 100.0%; Pred. No. 9.2e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 741 tcggaattcgagctcgctgcacaagcttgcgcccagctcgagacaccaccaccacca 800
|||||
1715 TCCGAATTCGAGCTCGCTGCACAGCTTGCGCCGCACTCGACGACACACACACACCA 1656

QY 801 ctga 804
|||||
DB 1655 CTGA 1652

RESULT 2
AAx28085
ID AAX28085 standard; DNA; 5616 BP.
XX
AC AAX28085;

XX 10-JUN-1999 (first entry)

DE Plasmid pET21.tac.

XX
KM MuLV reverse transcriptase; dsCDNA production; cDNA library production;
ss.

XX Synthetic.

XX US5891637-A.

XX 06-APR-1999.

XX 15-SEP-1997; 97US-0929967.

XX 15-OCT-1996; 96US-0732861.

XX 15-SEP-1997; 97US-0929967.

XX (GETH) GENENTECH INC.

XX Ruppert SJ;

XX WPI; 1999-253852/21.

XX Producing double stranded cDNA molecules from mRNA transcripts and a
XX novel xenotropic murine leukemia virus reverse transcriptase -
XX useful for constructing cDNA libraries
XX
XX Example 5; Fig 7; 70pp; English.

CC This sequence represents a plasmid used to test the method of
CC the invention. The method is for the production of double
CC stranded cDNA (dsCDNA) molecules from mRNA transcripts and a xenotropic
CC murine leukemia virus reverse transcriptase (xm-MuLV). The method, the
CC host cells, reverse transcriptases and nucleic acids disclosed may be
CC used to produce high quality, normalized, full-length, directionally
CC cloned, cDNA libraries. Use of the method to construct cDNA libraries
CC ensures that the reverse transcriptase is able to extend the first strand of
CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
CC all sequences are represented in the library, whereas previously, some
CC sequences would be lost due to incomplete transcription (e.g. as a result
CC of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
CC H activity). Also, carrying the process out in cells means that enzymes
CC and proteins present in the cell will repair any mistakes or nicks in the
CC product DNA. Modifying the mRNA for insertion into a vector by adding the
CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
CC the method provides a fast and clean way of synthesizing cDNA
CC clones.
XX
SQ Sequence 5616 BP; 1306 A; 1478 C; 1492 G; 1340 T; 0 other;

Query Match 8.0%; Score 64; DB 20; Length 5616;
Best Local Similarity 100.0%; Pred. No. 9e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcacaagcttgcgcccagctcgagacaccaccaccacca 800
|||||
DB 5417 TCCGAATTCGAGCTCGCTGCACAGCTTGCGCCGCACTCGACGACACACACACCA 5476

QY 801 ctga 804
|||||
DB 5477 CTGA 5480

RESULT 3
AAx80369
ID AAX80369 standard; cDNA; 2007 BP.
XX
AC AAX80369;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex encoding cDNA SEQ ID:NO:119.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;

XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

XX hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

XX Synthetic.

XX WO928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes
XX
XX Disclosure; Page 204-207; 211pp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
XX
XX

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgctgcacaagcttgcggcgacactgcagaccacaccacacactg 803
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1942 gaattcgagctcgctgcacaagcttgcggcgacactgcagaccacaccacacactg 2001
 OY 804 a 804
 |
 Db 2002 a 2002

RESULT 6

AAx80367
 ID AAX80367 standard; cDNA: 2007 BP.

XX AAX80367;

DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:117.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;

KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

PN WO928482-A2.

XX 10-JUN-1999.

PD 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

PA Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI: 1999-385385/32.

DR WPI: 1999-385385/32.

XX New hepatitis C virus covalent complexes

PT New hepatitis C virus covalent complexes

XX Disclosure: Page 197-200; 211pp; English.

PS The present invention describes a covalent hepatitis C virus (HCV)

CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the

CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker

CC to the amino terminus of the HCV NS3 protease domain. The present

CC sequence encodes an example of the above complex. The covalent

CC NS4A-NS3 complexes are useful for structural determination and

CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

CC They can also be used for detecting inhibitors of the protease activity,

CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the non-

CC covalent protease-peptide complexes previously available.

CC Sequence 2007 BP: 415 A; 619 C; 564 G; 409 T; 0 other;

XX Query Match 7.6%; Score 61; DB 20; Length 2007;

XX Best Local Similarity 100.0%; Pred. No. 2.8e-20;

XX Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgctgcacaagcttgcggcgacactgcagaccacaccacacactg 803

Db 1942 gaattcgagctcgctgcacaagcttgcggcgacactgcagaccacaccacacactg 2001

OY 804 a 804

AC

XX

Db 2002 a 2002

RESULT 7

AAx80368
 ID AAX80368 standard; cDNA: 2007 BP.

XX AAX80368;

DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:118.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;

KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

PN WO928482-A2.

XX 10-JUN-1999.

PD 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

PA Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI: 1999-385385/32.

DR WPI: 1999-385385/32.

XX New hepatitis C virus covalent complexes

PT New hepatitis C virus covalent complexes

XX Disclosure: Page 201-203; 211pp; English.

PS The present invention describes a covalent hepatitis C virus (HCV)

CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the

CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker

CC to the amino terminus of the HCV NS3 protease domain. The present

CC sequence encodes an example of the above complex. The covalent

CC NS4A-NS3 complexes are useful for structural determination and

CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

CC They can also be used for detecting inhibitors of the protease activity,

CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the non-

CC covalent protease-peptide complexes previously available.

CC Sequence 2007 BP: 414 A; 620 C; 563 G; 410 T; 0 other;

XX Query Match 7.6%; Score 61; DB 20; Length 2007;

XX Best Local Similarity 100.0%; Pred. No. 2.8e-20;

XX Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgctgcacaagcttgcggcgacactgcagaccacaccacacactg 803

Db 1942 gaattcgagctcgctgcacaagcttgcggcgacactgcagaccacaccacacactg 2001

OY 804 a 804

Db 2002 a 2002

RESULT 8

AAx44351
 ID AAT44351 standard; cDNA: 2735 BP.

XX AAT44351;

AC

XX 11-FEB-1997 (first entry)
 DT
 XX H. pylori ureA + ureB gene locus in pORV214.
 DE
 XX Urease; urea gene; ureB gene; vaccine; ds.
 KW
 XX Chimeric Helicobacter strain CPM630;
 OS
 XX Chimeric bacteriophage T7.
 FH
 Key
 Location/Qualifiers
 FT 1.16
 FT /*tag= a
 FT /note= "T7 promoter provides transcription
 FT initiation for the urease genes"
 FT misc_signal
 FT 33..43
 FT /*tag= b
 FT /function= operator
 FT /note= "lac operator provides inducible expression
 FT of the urease genes"
 FT primer_bind
 FT complement (46..67)
 FT /*tag= c
 FT /note= "Bul primer"
 FT 102..818
 FT /*tag= d
 FT /product= urease A subunit
 FT 822..2531
 FT /*tag= e
 FT /product= urease B subunit
 FT 2546..2569
 FT primer_bind
 FT /*tag= f
 FT /note= "Bul2 primer"
 FT 2693..2735
 FT /*tag= g
 FT /note= "T7 terminator"
 FT terminator
 FT /note= "T7 terminator"
 XX WO9633732-A1.
 PN
 XX 31-OCT-1996.
 PD
 XX 25-APR-1996; 96WO-US056800.
 PE
 XX 06-DEC-1995; 95US-0568122.
 PR
 XX 28-APR-1995; 95US-0431041.
 PR
 XX (ORAV-) ORAVAX INC.
 PA
 XX Ackerman SK, Bhagat H, Ermak T, Guirakhoo F, Kleantous H;
 PI Lee CK, Monath TP, Pappo J, Soman G, Sussman I;
 PI Thomas WD, Weltzin RA;
 PI
 XX WPI: 1996-497373/49.
 DR P-PSDB; AAM07193;
 DR P-PSDB; AAM07194.
 DR
 XX Vaccine for inducing mucosal response to Helicobacter - contg.
 PT multimeric urease complex and pref. an antibiotic, anti-secretory
 PT agent or bismuth salt
 PT
 XX
 XX Disclosure: Page 68-70; 98pp; English.
 PS
 XX A cDNA clone (AAT44351), derived from pORV214, includes the ureA and
 CC ureB sequences coding for the urease A (AAM07193) and urease B
 CC (AAM07194) subunits of Helicobacter pylori clinical isolate CPM630.
 CC To obtain pORV214, a genomic DNA library of CPM630 was screened
 CC with anti-Helicobacter urease antibody. A 17 kb SalI fragment from
 CC an isolated clone was subcloned into pUC18 to give pSCPI. PCR
 CC primers (AAT44352-53) were used to amplify a 2.5 kb fragment from
 CC pSCPI, which was inserted into PET24+ to give pORV214. The vector
 CC was utilised in the produ. of recombinant, enzymatically inactive,
 CC multimeric urease in E. coli transformants for use in vaccines to
 CC treat or prevent Helicobacter infection.
 CC
 XX

SQ Sequence 2735 BP; 875 A; 562 C; 635 G; 663 T; 0 other;
 Query Match 7.6%; Score 61; DB 17; Length 2735;
 Best Local Similarity 100.0%; Pred. No. 2.8e-20;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 744 gaattcgagctcgacgaagcttgcgcccactcgagaccaccaccacactg 803
 Db 2564 gaattcgagctcgacgaagcttgcgcccactcgagaccaccaccacactg 2623
 QY 804 a 804
 Db 2624 a 2624
 RESULT 9
 AAT08489
 ID AAT08489 standard; DNA; 5248 BP.
 XX
 AC AAT08489;
 XX
 DT 15-MAR-1996 (first entry)
 XX
 DE Plasmid PET Trc S05/N1 containing 561.1 scFv.
 XX
 KW Complement C5; haemolysis; kidney; glomerulonephritis;
 KW monoclonal antibody; antiinflammatory; antibody engineering;
 KW humanised antibody; complementarity determining region; CDR;
 KW single chain antibody; scFv; vector; PET Trc S05/N1;
 KW Escherichia coli; ds.
 XX
 OS Synthetic.
 XX
 PN WO9529697-A1.
 PN
 XX 09-NOV-1995.
 PD
 XX 01-MAY-1995; 95WO-US05688.
 PE
 XX 02-MAY-1994; 94US-0236208.
 PR
 XX (ALEX-) ALEXION PHARM INC.
 PA
 XX Evans MJ, Malis L, Mueller EE, Nye SH, Rollins S;
 PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;
 PI Wang T, Wilkins JA;
 PI
 XX WPI: 1995-392923/50.
 DR
 XX Treating glomerulonephritis with antibody against complement C5
 PT component - to inhibit complement induced cell lysis
 PT
 XX
 XX Example 11; Page 141-146; 181pp; English.
 PS
 XX A DNA construct (AAT08488) coding for a humanised CDR-grafted
 CC scFv, designated 561.1 scFv D012 (AAR77616), was subcloned into
 CC bacterial expression vector PET Trc S05/N1 and the resulting
 CC plasmid (sequence given in AAT08489) was transformed into
 CC Escherichia coli W61. Humanised, recombinant scFv was obtd.
 CC which retained the ability of Mab 561.1 to block human complement
 CC C3a and C5b-9 generation and which was thus able to reduce
 CC glomerular inflammation and kidney dysfunction associated with
 CC glomerulonephritis.
 CC
 XX
 SQ Sequence 5248 BP; 1226 A; 1373 C; 1381 G; 1268 T; 0 other;
 Query Match 7.6%; Score 61; DB 16; Length 5248;
 Best Local Similarity 100.0%; Pred. No. 2.6e-20;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 744 gaattcgagctcgacgaagcttgcgcccactcgagaccaccaccacactg 803

DB 5052 gaattcgagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactg 5111
 QY 804 a 804
 DB 5112 a 5112

RESULT 10
 AAT41898
 ID AAT41898 standard; DNA: 5248 BP.

AC AAT41898;

DT 01-FEB-1997 (first entry)

DE Vector pET Trc SOS/NI.

XX Vector; PET Trc SOS/NI; myelin basic protein; MBP;
 KM proteolipid protein; PLP; multiple sclerosis; autoimmune disease;
 KM diagnosis; therapy; T-lymphocyte; T-cell; anergy; apoptosis; ds;
 KM cyclic.

OS Synthetic.

PN WO9634622-A1.

PD 07-NOV-1996.

PF 22-APR-1996; 96WO-US05611.

PR 07-JUN-1995; 95US-0482114.

PR 02-MAY-1995; 95US-0431644.

PR 02-MAY-1995; 95US-0431648.

XX (ALEX-) ALEXION PHARM INC.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Leonardo MJ, Matis L, McFarland HP, Mueller EE, Mueller JP;

PI Nye SH, Pelfrey CM, Squinto SP, Wilkins JA;

PI WPI; 1996-505898/50.

PS New human myelin basic protein and proteolipid protein variant(s)

CC used in the assessment, diagnosis and treatment of multiple

CC sclerosis

CC Disclosure: Page 100-103, 156pp; English.

CC Plasmid PET Trc SOS/NI (AAT41898) is a bacterial expression vector.

CC It is suitable for expression of human native and modified myelin

CC basic protein and proteolipid protein (see also AAM00399-400 and

CC AAM06101-08) in bacterial host cells, e.g. Escherichia coli.

CC Sequence 5248 BP; 1226 A; 1373 C; 1381 G; 1268 T; 0 other;

QY 744 gaattcgagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactg 803
 DB 5052 gaattcgagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactg 5111

QY 804 a 804
 DB 5112 a 5112

RESULT 11
 AAX28091
 ID AAX28091 standard; DNA: 6363 BP.

XX AAX28091;
 AC 10-JUN-1999 (first entry)
 XX
 DT Plasmid pACYC-pol.
 DE
 XX
 KM MULV reverse transcriptase; dsDNA production; cDNA library production;
 KM ss.

OS Synthetic.
 XX US5891637-A.
 PN
 PD 06-APR-1999.

PF 15-SEP-1997; 97US-0929967.

PR 15-OCT-1996; 96US-0732861.

PR 15-SEP-1997; 97US-0929967.

XX (GETH) GENENTECH INC.

PA Ruppert SJ;

PI WPI; 1999-253852/21.

DR Producing double stranded cDNA molecules from mRNA transcripts and a

PT novel xenotropic murine leukemia virus reverse transcriptase -

PT useful for constructing cDNA libraries

XX Example 5; Fig 11; 70pp; English.

PS This sequence represents a plasmid used to test the method of

CC the invention. The method is for the production of double

CC stranded cDNA (dsDNA) molecules from mRNA transcripts and a xenotropic

CC murine leukaemia virus reverse transcriptase (XV-MuLV). The method, the

CC host cells, reverse transcriptases and nucleic acids disclosed may be

CC used to produce high quality, normalized, full-length, directionally

CC cloned, cDNA libraries. Use of the method to construct cDNA libraries

CC ensures that the reverse transcriptase is able extend the first strand of

CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that

CC all sequences are represented in the library, whereas previously, some

CC sequences would be lost due to incomplete transcription (e.g. as a result

CC of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase

CC H activity). Also, carrying the process out in cells means that enzymes

CC and proteins present in the cell will repair any mistakes or nicks in the

CC product DNA. Modifying the mRNA for insertion into a vector by adding the

CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,

CC the method provides a fast and clean way of synthesizing cDNA

CC clones.

CC Sequence 6363 BP; 1576 A; 1701 C; 1645 G; 1441 T; 0 other;

QY 750 gagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactga 804
 DB 4656 gagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccaccactga 4710

RESULT 12
 AAF74635
 ID AAF74635 standard; DNA: 59 BP.
 AC AAF74635;
 XX
 DT 15-MAY-2001 (first entry)
 DE Cloning vector pBK77 nucleotide fragment SEQ ID NO:9.

XX Nicking endonuclease; N.BstNBI; restriction endonuclease;
 XX non-thio strand displacement amplification; P1el modification methylase;
 KW ds.
 XX
 XX Bacillus steaerothermophilus.
 OS
 XX
 XX US6191267-B1.
 PN
 XX
 XX 20-FEB-2001.
 PD
 XX
 XX 02-JUN-2000; 2000US-0586935.
 PF
 XX
 XX 02-JUN-2000; 2000US-0586935.
 PR
 XX
 XX (NEME) NEW ENGLAND BIOLABS INC.
 PA
 XX
 XX Kong H, Higgins LS, Dalton M, Kucera RB, Schildkraut I;
 PI
 XX
 XX WPI; 2001-210381/21.
 DR
 XX
 XX Recombinant DNA encoding a novel nicking endonuclease, N.BstNBI useful
 PT in non-thio strand displacement amplification -
 PR
 XX
 XX Disclosure; Fig 6; 30pp; English.
 PS
 XX
 XX The present invention describes an isolated DNA (I) coding for the
 CC N.BstNBI restriction endonuclease obtainable from ATCC Accession
 CC No. PTA-1925. Also described are: (1) a vector (II) comprising (I);
 CC (2) a host cell transformed by (II); and (3) producing an N.BstNBI
 CC restriction endonuclease by culturing a host cell transformed with
 CC (II) under conditions suitable for expression of the endonuclease.
 CC (I) is useful for producing N.BstNBI nicking endonuclease, which
 CC is useful in non-thio strand displacement amplification. The present
 CC sequence represents a fragment from the pKRT7 cloning vector
 CC which is given in the exemplification of the present invention.
 CC
 CC Sequence 59 BP; 11 A; 22 C; 17 G; 9 T; 0 other;

	Query March	6.2%; Score 50; DB 22; Length 59;
	Best Local Similarity	100.0%; Pred. No. 8 3e-15;
	Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
OY	741 tccgaattcgagctccgctgcacaagaacttgcggcgcacatcgagcaacc	790
Db	10 tccgaattcgagctccgctgcacaagaacttgcggcgcacatcgagcaacc	59
	RESULT 13	
	AAA61167	
ID	AAA61167 standard; DNA; 357 BP.	
XX		
AC	AAA61167;	
XX		
DT	12-OCT-2000 (first entry)	
DE		
XX		
SEN	virus genome fragment SEQ ID NO: 137.	
DE		
KW	SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease; proliferative disorder; hepatopathy; hepatitis; viral infection;	
KM	vacination; gene therapy; ds.	
OS	Hepatitis virus.	
XX		
PX	WC200028039-A2.	
PN		
XX		
PD	18-MAY-2000.	
XX		
PE	09-NOV-1999; 99WO-EP08566.	
XX		
XX	10-NOV-1998; 98IT-MI02437.	
RR	30-APR-1999; 99IT-MI00923.	

PR	14-MAY-1999;	99EP-0830298.
PR	16-JUL-1999;	99EP-0113932.
XX		
PA	(DIAS-) DIASORIN SRL.	
PI	Primi D, Fioridalisi G, Mantero GL, Mattioli S, Sottini A;	
PI	Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;	
XX		
DR	WPI; 2000-376551/32.	
XX		
PT	Nucleic acids representing the genome of the SEN virus (SENV) and	
PT	encoded proteins, useful for treatment of hepatopathies, inflammatory	
PT	diseases and proliferative disorders such as cancer -	
XX		
PS	Example 25; Page 97; 392pp; English.	
CC		
CC	The present invention is concerned with the sequence of the genome of the	
CC	SEN virus (SENV), and the proteins encoded by it. SENV is thought to be	
CC	the cause of hepatopathies which are not linked to the presence of the	
CC	hepatitis A, B and E viruses in man. The genome and proteins of this	
CC	virus can be used in gene therapy and vaccination against the virus,	
CC	which also causes disorders of the gastrointestinal tract, including	
CC	Crohn's disease and lupus erythematosus, inflammatory diseases, and	
CC	proliferative disorders such as cancer.	
XX		
SO	Sequence 357 BP; 117 A; 109 C; 76 G; 55 T; 0 other;	

Query Match	5.2%; Score 42;	DB 21;	Length 357;
Best Local Similarity	100.0%;	Pred. No. 5,9e-11;	
Matches	42;	Conservative	0; Mismatches: 0; Indels 0; Gaps

Oy	763	aagcttgagcgcgactcgagcaccacccacccactga	804
Db	316	aagcttgagcgcgactcgagcaccacccacccactga	357

RESULT	14
AAA61166	
ID	AAA61166 standard; DNA; 586 BP.
XX	
XX	AAA61166;
XX	
DT	12-OCT-2000 (first entry)
XX	
DE	SEN virus genome fragment SEQ ID NO: 133.
XX	
KM	SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
KM	proliferative disorder; hepatopathy; hepatitis; viral infection;
KW	vaccination; gene therapy; ds.
XX	
OS	Hepatitis virus.
XX	
PN	WO200028039-A2.
XX	
PD	18-MAY-2000.
XX	
XX	
PF	09-NOV-1999; 99WO-EP08566.
XX	
XX	
PR	10-NOV-1998; 98IT-MI02437.
PR	30-APR-1999; 99IT-MI00923.
PR	14-MAY-1999; 99EP-0830298.
PR	16-JUL-1999; 99EP-0113932.
XX	
PA	(DIAS-) DIASORIN SRL.
XX	
PI	Primi D, Fioridalisi G, Mantero GL, Mattioli S, Sottini A;
PI	Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
XX	
DR	WPI; 2000-376551/32.

Nucleic acids representing the genome of the SEN virus (SENV) and encoded proteins, useful for treatment of hepatopathies, inflammatory

PT diseases and proliferative disorders such as cancer -

PS Example 24; Page 94; 392pp; English.

CC The present invention is concerned with the sequence of the genome of the
CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
CC the cause of hepatopathies which are not linked to the presence of the
CC hepatitis A, B and E viruses in man. The genome and proteins of this
CC virus can be used in gene therapy and vaccination against the virus,
CC which also causes disorders of the gastrointestinal tract, including
CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
CC proliferative disorders such as cancer.

SO Sequence 586 BP; 155 A; 172 C; 139 G; 119 T; 1 other;

Query Match 5.2%; Score 42; DB 21; Length 586;

Best Local Similarity 100.0%; Pred. No. 5.7e-11;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aagcttgcgagcgagcgcagcagccaccaccaccactga 804
|||||
DB 545 aagcttgcgagcgagcgcagcagccaccaccaccactga 586

RESULT 15

ID AAX29726 standard; DNA; 1186 BP.

AC AAX29726;

DT 08-JUN-1999 (first entry)

DE Insert from pET21C-clone 2 from MSRV-1.

KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KM rheumatoid polyarthritis; ss.

OS Multiple sclerosis related virus type 1.

PN FR2765588-A1.

PD 08-JAN-1999.

PF 07-JUL-1997; 97FR-0008816.

PR 07-JUL-1997; 97FR-0008816.

PA (INMR) BIO MERIEUX.

DR WPI; 1999-098275/09.

DR P-PsDB; AAW99558.

PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis

PS Disclosure: Fig 8; 83pp; French.

CC This sequence represents the insert found in the plasmid pET21C-clone 2
CC containing a fragment from a novel multiple sclerosis related virus
CC type 1 (MSRV1). The sequence can be used in diagnostic, prophylactic or
CC therapeutic compositions to inhibit expression of a multiple sclerosis
CC related virus and/or virus associated with rheumatoid polyarthritis.

SO Sequence 1186 BP; 378 A; 284 C; 260 G; 264 T; 0 other;

Query Match 5.2%; Score 42; DB 20; Length 1186;

Best Local Similarity 100.0%; Pred. No. 5.5e-11;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aagcttgcgagcgagcgcagcagccaccaccaccactga 804
|||||

DB 1096 aagcttgcgagcgagcgcagcagccaccaccaccaccactga 1137

Search completed: August 23, 2001, 15:01:33
Job time: 9851 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 14:09:06 ; Search time 1867.27 Seconds
(without alignments)
4070.164 Million cell updates/sec

Title: us-09-654-652a-5

Perfect score: 804
Sequence: 1 atggttagcgcaaggattt.....accaccaccaccaccactga 804

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size: 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	2.7	154	153	BG412308	BG412308 OV2_40.D0
2	22	2.7	163	173	BG054383	BG054383 OV2_3-G03
3	22	2.7	290	148	BF20855	BF20855 OV1_12-A0
4	22	2.7	380	114	AM287483	AM287483 LG1_228-E
5	22	2.7	476	152	BG322490	BG322490 EML_28-F0
6	22	2.7	509	104	AT948373	AT948373 603044A01
7	22	2.7	510	156	D15218	D15218 RICC0286A R
8	22	2.7	544	137	BE594445	BE594445 P11_33-E0
9	22	2.7	548	154	BG487461	BG487461 EML_65-C1
10	22	2.7	555	154	BG487511	BG487511 EML_65-H1
11	22	2.7	581	137	BE598959	BE598959 P11_84-D0
12	22	2.7	584	137	BE600643	BE600643 P11_89-F0
13	22	2.7	615	137	BE598799	BE598799 P11_82-A1
14	22	2.7	618	166	BE357292	BE357292 DGI_148-C
15	21	2.6	272	155	BG550683	BG550683 sad22f12.
16	21	2.6	272	160	BB537820	BB537820 BB537820
17	21	2.6	292	162	BE000537	BE000537 RC3-BN007
18	21	2.6	300	151	BE597387	BE597387 su97b10.Y
19	21	2.6	309	160	BB540994	BB540994 BB540994
20	21	2.6	311	128	BB218522	BB218522 BB218522
21	21	2.6	337	31	AV630939	AV630939 AV630939
22	21	2.6	385	121	AM838442	AM838442 CM0-17005
23	21	2.6	394	158	HA9830	HA9830 yq21908.r1
24	21	2.6	406	30	AV426195	AV426195 AV426195
25	21	2.6	413	22	AT619419	AT619419 CEST2000
26	21	2.6	430	244	AA7472019	AA7472019 IM0286020
27	21	2.6	440	107	AU015444	AU015444 AU015444
28	21	2.6	441	219	CNS01P0U	AL155086 Anopheles
29	21	2.6	473	244	AZ470822	AZ470822 IM0355C13
30	21	2.6	480	245	AZ498394	AZ498394 IM0335J10
31	21	2.6	483	250	AZ853499	AZ853499 ZM0156A16
32	21	2.6	487	107	AU084357	AU084357 AU084357
33	21	2.6	494	231	AO655902	AO655902 Sheared D
34	21	2.6	499	243	AZ392169	AZ392169 IM0154K15
35	21	2.6	507	164	BE212984	BE212984 IPBn01K33
36	21	2.6	534	241	AZ287683	AZ287683 RPT-23-1
37	21	2.6	546	241	AQ439343	AQ439343 HS-1060-B
38	21	2.6	556	228	AZ642854	AZ642854 IM0506M06
39	21	2.6	560	240	AZ261227	AZ261227 RPT-23-4
40	21	2.6	614	111	AM077378	AM077378 f134d10.Y
41	21	2.6	633	238	AZ114598	AZ114598 RPT-23-4
42	21	2.6	641	136	BE469249	BE469249 IPHAK0151
43	21	2.6	651	242	AZ366869	AZ366869 IM0116J15
44	21	2.6	678	241	AZ328225	AZ328225 IM0051F24
45	21	2.6	685	156	C94209	C94209 C94209 Dict

ALIGNMENTS

RESULT 1
 LOCUS BG412308 154 bp mRNA EST 13-MAR-2001
 DEFINITION OV2_40.D02.bl_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA
 sequence.
 ACCESSION BG412308
 VERSION BG412308.1 GI:13317861
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 154)
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 TITLE An EST database from Sorghum: ovaries of varying immature stages

JOURNAL COMMENT

Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 142
 POLYA-No.

FEATURES

Location/Qualifiers
 1..154
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Ovary 2 (OV2)"
 /note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 30 a 50 c 60 g 14 t

BASE COUNT

Query Match 2.7%; Score 22; DB 153; Length 154;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY

780 cgagcaccacccacccacac 801
 |||||||
 Db 65 CGAGCACCCACCCACCCAC 86

RESULT 2

LOCUS BG054383 163 bp mRNA EST 25-JAN-2001
 DEFINITION OV2_3.G03.bl_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA
 sequence.
 ACCESSION BG054383
 VERSION BG054383.1 GI:12511071
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 163)
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 TITLE An EST database from Sorghum: ovaries of varying immature stages
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 160
 POLYA-No.

FEATURES

Location/Qualifiers
 1..163
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Ovary 2 (OV2)"
 /note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 30 a 61 c 54 g 18 t

ORIGIN

Query Match 2.7%: Score 22; DB 173; Length 163;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccaccaccaccaccac 801
|||||
Db 93 CGAGCACACCACCACCACCAC 114

RESULT 3

BF420855 290 bp mRNA EST 28-NOV-2000
LOCUS OVL_12_A02_b1_A002 Ovary 1 (OVL) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.

ACCESSION BF420855
VERSION BF420855.1 GI:11408844

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoidae; Andropogoneae; Sorghum.
1 (bases 1 to 290)

REFERENCE Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt

,L.H.
An EST database from Sorghum: ovaries of varying immature stages

Unpublished (2000)

CONTACT: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 140

POLYA-No.

FEATURES

source

1. 290

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_11b="Ovary 1 (OVL)"

/note="Organ: Mix of ovaries of varying immature stages

from 8-week-old plants; Vector: pBluescript II from lambda

zap II; Site_1: XhoI; Site_2: EcoRI; The library was made

from poly-A RNA in the cloning vector lambda zap II.

Clones to be sequenced were prepared by mass excision."

BASE COUNT 54 a 107 c 85 g 42 t 2 others

ORIGIN

Query Match 2.7%: Score 22; DB 148; Length 290;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccaccaccaccaccac 801

|||||

Db 104 CGAGCACACCACCACCACCAC 125

RESULT 4

AM287483 380 bp mRNA EST 19-JUL-2000

LOCUS LGL_228_E07_b1_A002 Light Grown 1 (LGL) Sorghum bicolor cDNA, mRNA

DEFINITION sequence.

ACCESSION AM287483

VERSION AM287483.2 GI:6859477

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoidae; Andropogoneae; Sorghum.
1 (bases 1 to 380)

REFERENCE Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.

An EST database from Sorghum: light-grown seedlings

Unpublished (2000)

On Jan 6, 2000 this sequence version replaced gi:6677327.

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 379

POLYA-No.

FEATURES

source

1. 380

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_11b="Light Grown 1 (LGL)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse)

seedlings; Vector: lambda zap; Site_1: XhoI; Site_2: EcoRI

; The library was made from poly-A RNA in the cloning

vector lambda zap II. Clones to be sequenced were

prepared by mass excision."

BASE COUNT 79 a 145 c 104 g 52 t

ORIGIN

Query Match 2.7%: Score 22; DB 114; Length 380;

Best Local Similarity 100.0%; Pred. No. 0.46;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccaccaccaccaccac 801

|||||

Db 110 CGAGCACACCACCACCACCAC 131

RESULT 5

BG322490 476 bp mRNA EST 27-FEB-2001

LOCUS EML_28_F03_b1_A002 Embryo 1 (EML) Sorghum bicolor cDNA, mRNA

DEFINITION sequence.

ACCESSION BG322490

VERSION BG322490.1 GI:13152168

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

Clade; Panicoidae; Andropogoneae; Sorghum.
1 (bases 1 to 476)

REFERENCE Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.

An EST database from Sorghum: developing embryos

Unpublished (2000)

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.
Seq primer: JEN REV
High quality sequence stop: 393
POLYA-No.

FEATURES
source

location/Qualifiers
1. 476
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 99 a 177 c 127 g 73 t
ORIGIN

Query Match 2.7%; Score 22; DB 152; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccaccaccaccac 801
|||||
Db 104 CGAGCACCCACCCACCCAC 125

RESULT 6
AI948373/c 509 bp mRNA EST 19-AUG-1999
LOCUS 603044A01.x1 603 - stressed root cDNA library from Wang/Bohnert lab
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION AI948373
VERSION AI948373.1 GI:5740683
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 509)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 603044 row: A column: 01.
FEATURES
source

location/Qualifiers
1. 509
/organism="Zea mays"
/cultivar="873"
/db_xref="taxon:4577"
/clone_lib="603 - stressed root cDNA library from
Wang/Bohnert lab"
/tissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E. coli XL Gold"
/note="Organ: root; Vector: pBluescriptII SK(+) XE;
Seedling stressed root cDNA library from Wang/Bohnert lab"
BASE COUNT 122 a 111 c 131 g 145 t
ORIGIN

Query Match 2.7%; Score 22; DB 104; Length 509;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccaccaccaccac 801
|||||
Db 248 CGAGCACCCACCCACCCAC 227

RESULT 7
DI5218/c 510 bp mRNA EST 08-JUL-1999
LOCUS RICC0286A Rice callus Oryza sativa subsp. japonica cDNA clone
DEFINITION RICC0286A, mRNA sequence.
ACCESSION DI5218
VERSION DI5218.1 GI:286410
KEYWORDS EST.
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euphorbiaceae; Oryzaeae; Oryza.
REFERENCE 1 (bases 1 to 510)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp
PROJECT "RGP".
Seq primer: oligo(dT).
FEATURES
source

location/Qualifiers
1. 510
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C0286A"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site 1: SalI; Site 2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 80 a 121 c 171 g 106 t 32 others
ORIGIN

Query Match 2.7%; Score 22; DB 156; Length 510;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccaccaccaccac 801
|||||
Db 174 CGAGCACCCACCCACCCAC 153

RESULT 8
BE594445 544 bp mRNA EST 18-AUG-2000
LOCUS P11_31_E02.D1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION BE594445
VERSION BE594445.1 GI:9849518
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 544)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
L.H.

TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 481
POLYA-No.

FEATURES
source Location/Qualifiers

1..548
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (PI1)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; Two-week-old sorghum plants (BFX 623 cultivar) were infected with pathogen (isolate FRM41 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 123 a 197 c 141 g 83 t
ORIGIN

Query Match 2.7%; Score 22; DB 137; Length 544;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 cgaagcaccacacacacacacac 801
|||||
DB 93 CGAGCACACACACACACACAC 114

RESULT 9
BG487461 548 bp mRNA EST 27-MAR-2001
LOCUS EML_65.C11.b1_A002 Embryo 1 (EM1) Sorghum bicolor CDNA, mRNA
DEFINITION
ACCESSION BG487461
VERSION BG487461.1 GI:13468696
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade: Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 548)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)

AUTHORS
TITLE
JOURNAL
COMMENT

The University of Georgia
Department of Botany
Contact: Cordonnier-Pratt MM
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 546
POLYA-No.

FEATURES
source Location/Qualifiers

1..548
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 121 a 194 c 137 g 96 t
ORIGIN

Query Match 2.7%; Score 22; DB 154; Length 548;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 cgaagcaccacacacacacacac 801
|||||
DB 31 CGAGCACACACACACACACAC 52

RESULT 10
BG487511 555 bp mRNA EST 27-MAR-2001
LOCUS EML_65.H11.b1_A002 Embryo 1 (EM1) Sorghum bicolor CDNA, mRNA
DEFINITION
ACCESSION BG487511
VERSION BG487511.1 GI:13468746
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade: Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 555)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)

AUTHORS
TITLE
JOURNAL
COMMENT

The University of Georgia
Department of Botany
Contact: Cordonnier-Pratt MM
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 476
POLYA-No.

FEATURES
source Location/Qualifiers

1..555
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 123 a 203 c 142 g 87 t
ORIGIN

Query Match 2.7%; Score 22; DB 154; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccacacacacacacac 801
|||||

Db 104 CGAGCACCACACACACACAC 125

RESULT 11
BE598959 581 bp mRNA EST 18-AUG-2000
LOCUS P11_84_D06.bl_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION BE598959
VERSION BE598959.1 GI:9854018
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 581)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@rttenga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 542
POLYA-No.

FEATURES
source Location/Qualifiers
1..581
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone.lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; Two-week-old sorghum plants (BRX 623
cultivar) were infected with pathogen (isolate FRM421 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector Lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."

BASE COUNT 125 a 209 c 145 g 102 t

ORIGIN

Query Match 2.7%; Score 22; DB 137; Length 581;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccacacacacacacac 801
|||||

Db 23 CGAGCACCACACACACACAC 44

RESULT 12
BE600643 584 bp mRNA EST 18-AUG-2000
LOCUS P11_89_F08.bl_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION BE600643
VERSION BE600643.1 GI:9855716
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 584)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@rttenga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 568
POLYA-No.

FEATURES
source Location/Qualifiers
1..584
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone.lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; Two-week-old sorghum plants (BRX 623
cultivar) were infected with pathogen (isolate FRM421 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector Lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."

BASE COUNT 130 a 211 c 149 g 94 t

ORIGIN

Query Match 2.7%; Score 22; DB 137; Length 584;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgagcaccacacacacacacac 801
|||||

Db 93 CGAGCACCACACACACACAC 114

RESULT 13
BE598799 615 bp mRNA EST 18-AUG-2000
LOCUS P11_82_A10.bl_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
DEFINITION

mrna sequence.
 BE598799
 BE598799.1 GI:9853872
 EST.
 SOURCE
 ORGANISM
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.
 1 (bases 1 to 615)
 Cordonier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt L.H.
 An EST database from Sorghum: pathogen-induced plants
 Unpublished (2000)
 Contact: Cordonier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: JEN REV
 High quality sequence stop: 587
 POLYA=NO.
 FEATURES
 source
 Location/Qualifiers
 1..615
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (PI1)"
 /note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from lambda zap II; Site.1: XhoI; Site.2: EcoRI; Two-week-old sorghum plants (BRX 623 cultivar) were infected with pathogen (isolate PM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
 BASE COUNT
 134 a 214 c 164 g 103 t
 ORIGIN
 Query Match 2.7%; Score 22; DB 137; Length 615;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 780 cgagcaccaccaccaccac 801
 ||||||||||||||||||||
 Db 69 CGAGCACACACACACACAC 90
 RESULT 14
 BE357292 618 bp mrna EST 20-JUL-2000
 LOCUS DGI_148.C02.b1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
 DEFINITION
 BE357292
 BE357292
 BE357292.1 GI:92298849
 EST
 SOURCE
 ORGANISM
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.
 1 (bases 1 to 618)
 Cordonier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt L.H.
 An EST database from Sorghum: dark-grown seedlings
 Unpublished (2000)
 Contact: Cordonier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: JEN REV
 High quality sequence stop: 617
 POLYA=NO.
 FEATURES
 source
 Location/Qualifiers
 1..618
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DGI)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector: lambda zap; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT
 134 a 222 c 157 g 105 t
 ORIGIN
 Query Match 2.7%; Score 22; DB 166; Length 618;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 780 cgagcaccaccaccaccac 801
 ||||||||||||||||||||
 Db 99 CGAGCACACACACACACAC 120
 RESULT 15
 BG550683 272 bp mrna EST 09-APR-2001
 LOCUS sad22f12.y1.Gm-cl074 Glycine max cDNA clone GENOME SYSTEMS CLONE
 DEFINITION
 ID: Gm-cl074-1055 5', mRNA sequence.
 BG550683
 BG550683.1 GI:13562463
 EST.
 SOURCE
 ORGANISM
 soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 272)
 Shoemaker,R., Kelm,P., Vodkin,L., Erpelding,J., Corvelli,V., Khanna A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thelising,B., Allen,M., Bowers Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information

call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Putative full length read
 vector to vector length is 273.

FEATURES

source

1..272
 location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl074-1055"
 /clone_lib="Gm-cl074"
 /tissue_type="seedlings induced for HR (hypersensitive
 response)"
 /dev_stage="9-11 day old"
 /lab_host="DH10B"
 /note="Vector: Bluescript II SK+; Site.1: EcoRI; Site.2:
 XhoI; The cDNA library was constructed from mRNA isolated
 from 9-11 day old seedlings that were induced for HR
 (hypersensitive response) by vacuum infiltrating plant
 tissue with *Pseudomonas syringae* pv. *glycinea* carrying the
 avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
 unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
 and 53 hrs after inoculation and their mRNA pooled equally
 for cDNA construction. The library was prepared using the
 Stratagene Bluescript II SK(+) library construction kit.
 Complementary DNA was synthesized from mRNA using a primer
 consisting of a poly(dT) sequence with an XhoI restriction
 site. EcoRI adaptors were ligated to the blunt-ended cDNA
 fragments followed by XhoI digestion. The cDNA insert is
 protected from XhoI digestion via methylation during first
 strand synthesis. The cDNA fragments were directionally
 cloned into the EcoRI-XhoI restriction site of the
 Bluescript vector. The ligated cDNA fragments were
 transformed into *E. coli* Electromax DH10B host cells. Plant
 care, inoculations, and library construction were
 performed by Steve Clough (Lila Vodkin lab, University of
 Illinois)."

BASE COUNT 79 a 58 c 55 g 80 t
 ORIGIN

Query Match 2.68; Score 21; DB 155; Length 272;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 783 gcaccaccaccaccacactg 803
 ||||||||||||||||||||
 Db 11 GCACCACCACCACCACCTG 31

Search completed: August 23, 2001, 14:09:23
 Job time: 8655 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:18:28 ; Search time 19.47 Seconds

(without alignments)
614.030 Million cell updates/sec

Title: US-09-654-652A-3

Perfect score: 349
Sequence: 1 MNIKRTAVKSAALAVAAAAA.....AKGAKVNPNGHKRYRVEFH 349

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	349	100.0	349	1 GUB_FIBSU	P17989 fibrobacter
2	337	2.6	337	1 HAIR_DROME	P14003 drosophila
3	9	2.6	537	1 MMSA_BOVIN	Q07536 bos taurus
4	9	2.6	918	1 KPCM_MOUSE	Q62101 mus musculus
5	9	2.6	1475	1 N153_HUMAN	P49790 homo sapien
6	8	2.3	37	1 ANP3_PSEAM	P02733 pseudopleur
7	8	2.3	82	1 ANPA_PSEAM	P04002 pseudopleur
8	8	2.3	268	1 ZEST_DROME	Q24597 drosophila
9	8	2.3	317	1 CYSD_RHIME	P36892 rhizobium m
10	8	2.3	317	1 YE80_MYCTU	P71761 mycobacteri
11	8	2.3	368	1 CHEB_PSEAE	Q087125 pseudomonas
12	8	2.3	387	1 SOX1_HUMAN	O005725 homo sapien
13	8	2.3	391	1 HME1_HUMAN	O05925 homo sapien
14	8	2.3	391	1 SOX1_MOUSE	P53783 mus musculu
15	8	2.3	401	1 HME1_MOUSE	P09065 mus musculu
16	8	2.3	407	1 F16P_PEA	P46275 plisum sativ
17	8	2.3	424	1 ULB7_HCMVA	P16770 human cytom
18	8	2.3	428	1 FXB2_MOUSE	O64733 mus musculu
19	8	2.3	429	1 OORA_MYCTU	O10387 mycobacteri
20	8	2.3	440	1 DCO_DROME	O76324 drosophila
21	8	2.3	458	1 A2AD_HUMAN	P35369 homo sapien
22	8	2.3	459	1 IP3K_RAT	P17105 rattus norv
23	8	2.3	461	1 A2AC_HUMAN	P18825 homo sapien
24	8	2.3	461	1 IP3K_HUMAN	P23677 homo sapien
25	8	2.3	470	1 ESCA_DROME	P25932 drosophila
26	8	2.3	483	1 ELAV_DROME	P16914 drosophila
27	8	2.3	519	1 ELAV_DROVI	P23241 drosophila
28	8	2.3	528	1 TY3H_HUMAN	P07101 homo sapien
29	8	2.3	542	1 CH12_RH1OL	P29027 rhizopus ol
30	8	2.3	559	1 PHF1_MOUSE	O921b8 mus musculu
31	8	2.3	563	1 ARX_MOUSE	O35085 mus musculu
32	8	2.3	574	1 ZEST_DROME	P09956 drosophila
33	8	2.3	611	1 XINA_PSEFL	P14768 pseudomonas

34	8	2.3	618	1 ZEST_DROVI	Q24762 drosophila
35	8	2.3	652	1 FXO1_MOUSE	Q9160 mus musculu
36	8	2.3	655	1 FXO1_HUMAN	Q12778 homo sapien
37	8	2.3	666	1 HNF4_DROME	P49866 drosophila
38	8	2.3	668	1 PAB5_ARATH	O05196 arabidopsis
39	8	2.3	732	1 FOK_DROME	O01842 drosophila
40	8	2.3	741	1 ERF2_PICPI	P23637 pichia pinu
41	8	2.3	766	1 EYA_DROME	Q05201 drosophila
42	8	2.3	801	1 ZFY_HUMAN	P08048 homo sapien
43	8	2.3	828	1 NAPA_ECOLI	P33937 escherichia
44	8	2.3	880	1 BRCA_DROME	Q24206 drosophila
45	8	2.3	1077	1 HIES_DROME	Q02308 drosophila

ALIGNMENTS

RESULT	ID	GUB_FIBSU	STANDARD:	PRT:	349 AA.
AC	P17989:				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)				
DE	(1,3-1,4-BETA-D-GLUCAN-4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).				
OS	Fibrobacter succinogenes (Bacteroides succinogenes).				
OC	Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;				
CC	Fibrobacter.				
OX	NCBI_TaxID=833;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.				
RC	STRAIN-ISOLATE S85;				
RX	MEDLINE=90299807; PubMed=2193918;				
RA	Teacher R.W., Etile J.D.;				
RT	"DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";				
RL	J. Bacteriol. 172:3837-3841(1990).				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.				
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M33676; AAA24896.1; -				
DR	PIR: A44507; A44507.				
DR	HSSP; P23904; IAKK.				
DR	InterPro: IPR000757; -				
DR	Pfam: PF00722; Glyco_hydro_16; 1.				
DR	PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.				
KW	Hydrolase; Glycosidase; Signal; Repeat.				
FT	SIGNAL	1	27		
FT	CHAIN	28	349		
FT	ACT_SITE	79	79		BETA-GLUCANASE.
FT	ACT_SITE	83	83		NUCLEOPHILE (BY SIMILARITY).
FT	DOMAIN	271	307		PROTON DONOR (BY SIMILARITY).
FT	REPEAT	271	277		5 X 7 AA TANDEN REPEATS OF P-X-S-S-S-X.
FT	REPEAT	278	284		
FT	REPEAT	285	291		
FT	REPEAT	292	298		
FT	REPEAT	301	307		
SO	SEQUENCE	349 AA;	37737 MW;		16DCAF5BDEFC578A CRC64;

Query Match 100.0%; Score 349; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKAVKAGALVAAAALTTTNSAKDSGAEVLTLEEVQYKFKPARKMAAAGTSS 60
 DB 1 MNKKAVKAGALVAAAALTTTNSAKDSGAEVLTLEEVQYKFKPARKMAAAGTSS 60
 QY 61 SMFLYONGSEIADGPRVVEVDIEVLGNKPSFOSNITTTGAKAQKTSEKHNAVSPADQA 120
 DB 61 SMFLYONGSEIADGPRVVEVDIEVLGNKPSFOSNITTTGAKAQKTSEKHNAVSPADQA 120
 QY 121 FHYYGLEWNTYRYRYMTDGDGEVKTREGQVSNLTGTGLFPLNMLSSSAAMVQGFDESKL 180
 DB 121 FHYYGLEWNTYRYRYMTDGDGEVKTREGQVSNLTGTGLFPLNMLSSSAAMVQGFDESKL 180
 QY 181 PLFOFIMVWVYKYTTPGCGSGSDFLTMDTNDPFDGSRMGKGMDFPDGRRVDTLTKNI 240
 DB 181 PLFOFIMVWVYKYTTPGCGSGSDFLTMDTNDPFDGSRMGKGMDFPDGRRVDTLTKNI 240
 QY 241 YSRDGMILIALTRKQGSFNGQVPRDEPAPQSSSSAPASSSSVPAASSSSAFV 300
 DB 241 YSRDGMILIALTRKQGSFNGQVPRDEPAPQSSSSAPASSSSVPAASSSSAFV 300
 QY 301 PPSSSSTNTAIGHKRTTPAVAKAHEHRLVNAKAKVNPNGHKRRVNFEEH 349
 DB 301 PPSSSSTNTAIGHKRTTPAVAKAHEHRLVNAKAKVNPNGHKRRVNFEEH 349

RESULT 2
 HAIR_DROME STANDARD; PRT; 337 AA.
 AC P14003; Q9VSN8;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HAIRY PROTEIN.
 GN H OR CG6494.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_TaxID=227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=90059896; PubMed=2479541;
 RA Rushlow C.A., Hogan A., Pierchin S.M., Howe K.M., Iardelli M.,
 RA Ish-Horowicz D.;
 RT "The Drosophila hairy protein acts in both segmentation and bristle
 RT patterning and shows homology to N-myc.";
 RL EMBO J. 8:3095-3103(1989).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=1073132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazey R.G., Champe W., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Mostreft A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venler E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP WRPW MOTIF.
 RX MEDLINE=95094252; PubMed=8001118;
 RA Paroush Z., Finley R.L., Jr., Kidd T., Wainwright S.M., Ingham P.W.,
 RA Brent R., Ish-Horowicz D.;
 RT "Groucho is required for Drosophila neurogenesis, segmentation, and
 RT sex determination and interacts directly with hairy-related bHLH
 RT proteins.";
 RL Cell 79:805-815(1994).
 CC -1- FUNCTION: PAIR-RULE PROTEIN THAT REGULATES EMBRYONIC SEGMENTATION
 CC AND ADULT BRISTLE PATTERNING. TRANSCRIPTIONAL REPRESSOR OF GENES
 CC THAT REQUIRE A BHLH PROTEIN FOR THEIR TRANSCRIPTION (EG. THE FUSHI
 CC TARAZU GENE).
 CC -1- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
 CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
 CC HELIX-INTERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),
 CC RATHER THAN THE CANONICAL E-BOX (CANNGG).
 CC -1- DOMAIN: THE CARBOXY-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL
 CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO. A
 CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
 CC HAIRY-RELATED PROTEINS.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X15904; CAA34018.1; -
 CC EMBL: X15905; CAA34019.1; -
 CC EMBL: AEO03554; AAF50378.1; -
 CC PIR: S06956; S06956.
 CC TRANSFAC: T00345; -
 CC DR Flybase: FBgn0001168; h.
 CC DR InterPro: IPR001092; -
 CC DR InterPro: IPR003015; -
 CC DR Pfam: PF00010; HLH; 1.
 CC DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1
 KW Nuclear protein; developmental protein; pair-rule protein;
 KW DNA-binding; transcription regulation; repressor; polymorphism.
 FT DNA-BINDING 32 44
 FT DOMAIN 45 89 BASIC DOMAIN.
 FT DOMAIN 149 157 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 157 237 GIN-RICH.
 FT DOMAIN 222 237 GIN-RICH.
 FT DOMAIN 241 250 POLY-ALA.
 FT VARIANT 292 292 S -> P.
 FT DOMAIN 334 337 WRPW MOTIF (REQUIRED FOR ACTIVITY).
 SO SEQUENCE 337 AA; 36995 MW; 6D2ECAF7F2D56C0B CRC64;

Query Match 2.6%; Score 9; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
DB 238 LAVAAAAA 246

RESULT 3

MMSA_BOVIN STANDARD; PRT; 537 AA.
ID MMSA_BOVIN STANDARD; PRT; 537 AA.
AC 007536;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE [ACYLATING], MITOCHONDRIAL
DE PRECURSOR (EC 1.2.1.27) (MMSDH).
GN MMSDH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
OX NCBI_TaxID=9913;
RN [1]
RP TISSUE=Liver.
RX MEDLINE=93293905; PubMed=8514806;
RA Dethaite I., Berthiaume L., Pesceckis S.M., Patton W.F., Resh M.D.;
RT "Novel use of an iodo-myristyl-CoA analog identifies a semialdehyde
dehydrogenase in bovine liver."
RT J. Biol. Chem. 268:13738-13747(1993).
CC -1- FUNCTION: PLAYS A ROLE IN VALINE AND PYRIMIDINE METABOLISM. BINDS
FATTY ACYL-CoA.
CC -1- CATALYTIC ACTIVITY: 2-METHYL-3-OXOPROPANATE + COA + NAD(+) =
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- PPM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08643; AAA30650.1; -
DR PIR: A46600; A46600.
DR HSP: P56533; A4AS.
DR InterPro: IPR002086; -
DR Pfam: PF00171; aldehyd. 1.
DR PROSITE: PS00070; ALDEHYDE DEHYDR. CYS; 1.
DR PROSITE: PS00687; ALDEHYDE DEHYDR. GLU; FALSE NEG.
KW Oxidoreductase; NAD; Transit peptide; Mitochondrion.
FT TRANSIT 1 34
FT CHAIN 35 537
FT NP_BIND 263 268
FT ACT_SITE 319 319
FT CONFLICT 126 127
FT SEQUENCE 537 AA; 58062 MW; 69DE39506E2F9C0 CRC64;
ML -> TD (IN AA SEQUENCE).
BY SIMILARITY.
Query Match 2.6%; Score 9; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 AVAAAAAL 21

DB 5 AVAAAAAL 13

RESULT 4
KPCM_MOUSE STANDARD; PRT; 918 AA.

ID KPCM_MOUSE STANDARD; PRT; 918 AA.
AC 062101;
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE PROTEIN KINASE C, MU TYPE (EC 2.7.1.1) (NPKC-MU) (PROTEIN KINASE D).
GN PPKCM OR PKCM OR PKD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=94359973; PubMed=8078925;
RA Valverde A.M., Smet-Smit J., Van Lint J., Rozenburg E.;
RT "Molecular cloning and characterization of protein kinase D: a target
for diacylglycerol and phorbol esters with a distinctive catalytic
domain."
RT Proc. Natl. Acad. Sci. U.S.A. 91:8572-8576(1994).
CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
BINDING DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z34524; CAA84283.1; -
DR HSP: P28867; IPTQ.
DR MGD: MGI:99879; PKcm.
DR InterPro: IPR000719; -
DR InterPro: IPR001849; -
DR InterPro: IPR002219; -
DR InterPro: IPR002290; -
DR Pfam: PF00063; PKinase; 1.
DR Pfam: PF00130; DAG_PE-bind; 2.
DR Pfam: PF00169; PH; 1.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE: PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase; Serine/threonine-protein kinase;
KW Phorbol-ester binding; Zinc; Repeat.
FT DOMAIN 145 194
FT DOMAIN 277 326
FT DOMAIN 428 547
FT DOMAIN 589 845
FT DOMAIN 16 26
FT DOMAIN 198 201
FT NP_BIND 595 603
FT BINDING 618 618
FT ACT_SITE 712 712
ATP (BY SIMILARITY).
ATP (BY SIMILARITY).
ATP (BY SIMILARITY).
BY SIMILARITY.

SEQUENCE 918 AA; 102067 MW; 234486180521BDDBA CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 1; Length 918;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAAL 21
1111111111
Db 19 AVAAAAAAL 27

RESULT 5

ID N153_HUMAN STANDARD; PRT; 1475 AA.
AC P49730;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NUCLEAR PORE COMPLEX PROTEIN NUP153 (NUCLEOPORIN NUP153) (153 KDA NUCLEOPORIN).
GN NUP153.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=94154002; PubMed=8110839;
RA Morrow I., Bastos R., Horton H., Burke B.;
RT "Sequence analysis of a cDNA encoding a human nuclear pore complex protein, hnu153.";
RL Biochim. Biophys. Acta 1217:219-223(1994).
CC -1- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE TERMINAL RING STRUCTURE OF THE NUCLEOPLASMIC CAGE.
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST NUP1, NSP1, POM 121 AND MAMMALIAN P62.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z25535; CAA80982.1; -
CC MIM: 603948; -
DR InterPro: IPR001876; -
DR Pfam: PF00641; zf-RanBP; 4.
KW Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding.
FT DOMAIN 4 14
FT DOMAIN 443 447 POLY-RICH.
FT ZN_FING 662 681 C4-TYPE.
FT ZN_FING 726 745 C4-TYPE.
FT ZN_FING 797 816 C4-TYPE.
FT ZN_FING 855 874 C4-TYPE.
SQ SEQUENCE 1475 AA; 153889 MW; 3CB415A6909DF80E CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 1; Length 1475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 282 SSVPASSSS 290
1111111111
Db 823 SSVPASSSS 831

RESULT 6
ANP3_PSEAM STANDARD; PRT; 37 AA.
ID ANP3_PSEAM
AC P02733;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ANTIFREEZE PEPTIDE 3.
OS Pseudopleuronectes americanus (Winter flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes; OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP SEQUENCE
RX MEDLINE=78060969; PubMed=588591;
RA Davies A.L., Lin Y.;
RT "Structure of a peptide antifreeze and mechanism of adsorption to ice";
RL Biochim. Biophys. Acta 495:388-392(1977).
CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
DR PIR: A03192; PFEL3W.
DR HSSP: P04002; 1WFA.
DR InterPro: IPR00104; -
DR PRINTS: PR00308; ANTIFREEZE1.
KW Antifreeze protein; Repeat; Multigene family.
SQ SEQUENCE 37 AA; 3144 MW; 46AA951A962DECA9 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AVAAAAAL 22
1111111111
Db 6 AVAAAAAL 13

RESULT 7
ANP4_PSEAM STANDARD; PRT; 82 AA.
ID ANP4_PSEAM
AC P04002;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ANTIFREEZE PROTEIN A/B PRECURSOR.
OS Pseudopleuronectes americanus (Winter flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes; OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP SEQUENCE FROM N.A. (PROTEIN A).
RX MEDLINE=82197490; PubMed=6952188;
RA Davies P.L., Roach A.H., Hew C.-L.;
RT "DNA sequence coding for an antifreeze protein precursor from winter flounder";
RL Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
RN [2]
RP SEQUENCE FROM N.A. (PROTEIN A).
RX MEDLINE=88259236; PubMed=3133486;
RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
RT "Differential amplification of antifreeze protein genes in the pleuronectinae";
RL J. Mol. Evol. 27:29-35(1988).
RN [3]
RP SEQUENCE FROM N.A. (PROTEIN B).
RX MEDLINE=84264559; PubMed=6086629;

RA Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
 RT "Antifreeze protein genes of the winter flounder."
 RL J. Biol. Chem. 259:9241-9247(1984).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92209995; PubMed=1555765;
 RA Davies P.L.;
 RT "Conservation of antifreeze protein-encoding genes in tandem
 RL repeats."
 RN Gene 112:163-170(1992).
 OS [5]
 RP 3D-STRUCTURE MODELING OF 45-81.
 RX MEDLINE=92148833; PubMed=1738160;
 RA Chou K.-C.;
 RT "Energy-optimized structure of antifreeze protein and its binding
 RL mechanism."
 RN J. Mol. Biol. 223:509-517(1992).
 RP [6]
 RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
 RA Scherf F., Yang D.S.C.;
 RT "Ice-binding structure and mechanism of an antifreeze protein from
 RL winter flounder."
 RN Nature 375:427-431(1995).
 CC -I- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD-FREEZING POINT.
 CC -I- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
 CC -I- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL; L00138; AAB59964.1; -
 DR EMBL; L29178; AAB59964.1; JOINED.
 DR EMBL; M62414; AAA49469.1; -
 DR EMBL; X07506; CAA30389.1; -
 DR EMBL; M62416; AAA49471.1; -
 DR EMBL; M62417; AAA49472.1; -
 DR PIR; A03194; FDFLAW.
 DR PIR; A05161; A05161.
 DR PIR; S02326; S02326.
 DR PIR; JS0704; JS0704.
 DR PDB; 1ATE; 15-OCT-94.
 DR PDB; 1WFA; 03-JUN-95.
 DR PDB; 1WFB; 03-JUN-95.
 DR InterPro: IPR000104; -
 DR PRINTS: PR00308; ANTIFREEZE1.
 KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
 FT SIGNAL 1 21
 FT PROPEP 22 44
 FT
 FT CHAIN 45 82 REMOVED BY A DIPEPTIDYLPEPTIDASE
 FT VARIANT 36 36 (PROBABLY).
 FT VARIANT 70 70 ANTIFREEZE PROTEIN A/B.
 FT CONFLICT 24 24 A -> V.
 FT HELIX 46 80 A -> D (IN PROTEIN B).
 FT SEQUENCE 82 AA; 7711 MW; C2AE7B74C0D46CC1 CRC64; S -> R (IN REF. 2).
 Query Match 2.3%; Score 8; DB 1; Length 82;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
 ZEST DROMA STANDARD; PRT; 268 AA.
 ID ZEST DROMA 024597; 024598; 024599; 024600; 024601; 024602; 024603; 024604;
 AC 024597; 024598; 024599; 024600; 024601; 024602; 024603; 024604;
 AC 027387;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE REGULATORY PROTEIN ZESTE (FRAGMENT).
 GN 2.
 OS Drosophila mauritiana (Fruit fly),
 OS Drosophila sechellia (Fruit fly), and
 OS Drosophila simulans (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7226, 7238, 7240;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VARIOUS STRAINS;
 RX MEDLINE=93360802; PubMed=8355601;
 RA Hey J., Kilman R.M.;
 RT "Population genetics and phylogenetics of DNA sequence variation at
 RT multiple loci within the Drosophila melanogaster species complex."
 RL Mol. Biol. Evol. 10:804-822(1993).
 CC -I- FUNCTION: INVOLVED IN TRANSCRIPTION PHENOMENA (= SYNOPSIS-DEPENDENT
 CC GENE EXPRESSION), WHERE THE SYNAPTIC PAIRING OF CHROMOSOMES
 CC CARRYING GENES WITH WHICH ZESTE INTERACTS INFLUENCES THE
 CC EXPRESSION OF THESE GENES. ZESTE BINDS TO DNA AND STIMULATES
 CC TRANSCRIPTION FROM A NEARBY PROMOTER (BY SIMILARITY).
 CC -I- SUBUNIT: SELF-ASSOCIATE FORMING COMPLEXES OF SEVERAL HUNDRED
 CC MONOMERS (BY SIMILARITY).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL; L13049; AAA29032.1; -
 DR EMBL; L13050; AAA29033.1; -
 DR EMBL; L13051; AAA29034.1; -
 DR EMBL; L13052; AAA29035.1; -
 DR EMBL; L13053; AAA29036.1; -
 DR EMBL; L13054; AAA29037.1; -
 DR EMBL; L13055; AAA29038.1; -
 DR EMBL; L13056; AAA29039.1; -
 DR EMBL; L13057; AAA29040.1; -
 DR EMBL; L13058; AAA29041.1; -
 DR EMBL; L13059; AAA29042.1; -
 DR EMBL; L13060; AAA29043.1; -
 DR EMBL; L13061; AAA29044.1; -
 DR EMBL; L13062; AAA29045.1; -
 DR EMBL; L13063; AAA29046.1; -
 DR EMBL; L13064; AAA29047.1; -
 DR EMBL; L13065; AAA29048.1; -
 DR EMBL; L13066; AAA29049.1; -
 DR FlyBase: FBgn0012518; DmanuZ.
 DR FlyBase: FBgn0012801; DsecuZ.
 DR FlyBase: FBgn0012904; DsiniZ.
 KW DNA-binding; Transcription regulation; Nuclear protein.
 FT DNA_BIND 1 1
 FT NON_TER 1 1
 FT DOMAIN 97 268
 FT
 FT VARIANT 105 106 SPECIFIC, WITH ZESTE LOCUS.
 FT NON_TER 268 268 GUN/ALA-RICH (OPA-REPEAT INVOLVED IN
 FT TRANSCRIPTIONAL ACTIVATION OR REPRESSION
 FT AT DIFFERENT TARGET LOCI) (POTENTIAL).
 FT MISSING (IN STRAIN SI-K2).
 SO SEQUENCE 268 AA; 30353 MW; F592F8FA4F2DB47B CRC64;

Query Match 2.3%; Score 8; DB 1; Length 268;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAA 20
 Db 192 AVAAAAA 199

RESULT 9

CYS_D_RHIME STANDARD; PRT; 317 AA.
 AC P56892;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE SULFATE ADENYLYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLYLTRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT).
 GN CYS_D.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=582;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=99395034; PubMed=10464198;
 RA Abola A.P., Millits M.G., Wang R.C., Long S.R.;
 RT "Reduction of adenosine-5'-phosphosulfate instead of 3'-phosphodensosine-5'-phosphosulfate in cysteine biosynthesis by Rhizobium meliloti and other members of the family Rhizobiaceae";
 RL J. Bacteriol. 181:5280-5287(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Gallert F., Capela D., Hubler-Barloy F., Gattus M., Batut J., Boistard P., Gouzy J., Kahn D., Thebault P., Goffeau A., Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D., Vandenbol M., Puchler A., Becker A., Weidner S.;
 RA Submitted (MAY-2000) to the SWISS-PROT data bank;
 RL
 CC -1- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE + ADENYLYLSULFATE.
 CC -1- PATHWAY: FIRST STEP IN THE SULFATE ACTIVATION PATHWAY. THIS REACTION OCCURS EARLY IN THE REDUCTIVE BRANCH OF THE CYSTEINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF CYS_D, THE SMALLER SUBUNIT, AND CYSN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYS_D SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF158023; AAD55760.1; -
 DR InterPro: IPR002500; -
 DR Pfam: PF01507; PAPS_reduct; 1.
 KW Cysteine biosynthesis; Transferase; Nucleotidyltransferase.
 FT CONFLICT 260 260 R -> A (IN REF. 1).
 SO SEQUENCE 317 AA; 36429 MW; 093E2724D2841CB CRC64;

Query Match 2.3%; Score 8; DB 1; Length 317;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 243 RDGMILIA 250
 Db 234 RDGMILIA 241

RESULT 10
 YE80_MYCTU STANDARD; PRT; 317 AA.
 ID YE80_MYCTU
 AC P71761; 053171;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 34.3 KDA PROTEIN RV1480.
 GN RV1480 OR MTW007.27 OR MYCY2177.01.
 GN
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Jagsels K., Kirogh A., McLean J., Hamlin N., Holtroyd S., Hornsby T., Jagels K., Kirogh A., McLean J., Moulé S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence";
 RT Nature 393:537-544(1998).
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: STRONG, TO M. AVIUM MAV169.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL021184; CA116010.1; -
 DR Tuberculist: RV1480; -
 KW Hypothetical protein.
 SO SEQUENCE 317 AA; 34333 MW; 3D1D060FF5E5CE56 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 317;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 19
 Db 120 LAVAAAAA 127

RESULT 11

CHEB_PSEAE STANDARD; PRT; 368 AA.
 ID CHEB_PSEAE
 AC O87125;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PROTEIN-GLUTAMATE METHYLESTERASE (EC 3.1.1.61).
 GN CHEB OR PA1459.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PA01;
 RX MEDLINE=99161288; PubMed=10052136;
 RA Kato J., Nakamura T., Kuroda A., Ohtake H.;

RT "Cloning and characterization of chemotaxis genes in Pseudomonas
RL aeruginosa.";
RN Biosci. Biotechnol. Biochem. 63:155-161(1999).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy R., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
CC -!- FUNCTION: INVOLVED IN THE MODULATION OF THE CHEMOTAXIS SYSTEM;
CC CATALYZES THE DEMETHYLATION OF SPECIFIC METHYLGLUTAMATE RESIDUES
CC INTRODUCED INTO THE CHEMORECEPTORS (METHYL-ACCEPTING CHEMOTAXIS
CC PROTEINS) BY CHER (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PROTEIN L-GLUTAMATE O-METHYL ESTER + H(2)O =
CC PROTEIN L-GLUTAMATE + METHANOL.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- DOMAIN: THE N-TERMINAL REGULATORY DOMAIN INHIBITS THE ACTIVITY OF
CC THE C-TERMINAL EFFECTOR DOMAIN.
CC -!- PTM: PHOSPHORYLATED BY CHEA. PHOSPHORYLATION SUPPRESSES THE
CC INHIBITORY ACTIVITY OF THE N-TERMINAL DOMAIN (BY SIMILARITY).
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE RESPONSE
CC REGULATORY FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CHEB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AB012767; BAA33550.1; -;
CC DR EMBL; AE004575; AAG04848.1; -;
CC DR HSSP; P04042; ICHD.
CC DR InterPro; IPR000673; -;
CC DR InterPro; IPR001789; -;
CC DR Pfam; PF01339; CheB_methyltest; 1.
CC DR Pfam; PF00072; response_reg; 1.
CC KW Hydrolase; Chemotaxis; Sensory transduction; Phosphorylation.
FT DOMAIN 1 133 RESPONSE REGULATORY DOMAIN.
FT DOMAIN 134 177 LINKER.
FT DOMAIN 178 368 PROTEIN-GLUTAMATE METHYLESTERASE.
FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
FT ACT_SITE 192 192 BY SIMILARITY.
FT ACT_SITE 219 219 BY SIMILARITY.
FT ACT_SITE 312 312 BY SIMILARITY.
FT CONFLICT 34 34 G -> A (IN REF. 1).
SQ SEQUENCE -368 AA; 39004 MW; FE4801DC220C613B CRC64;

Query Match 2.3%; Score 8; DB 1; Length 368;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 APASSSSV 284
Db 151 APASSSSV 158
|||||||
RESULT 12
SOX1_HUMAN STANDARD; PRT; 387 AA.
AC O00570;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SOX-1 PROTEIN.
GN SOX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98051911; PubMed=9337405;
RA Malas S., Duthie S.M., Mohri F., Lovell-Badge R., Episkopou V.;
RA "Cloning and mapping of the human SOX1: a highly conserved gene
RT expressed in the developing brain.";
RL Mamm. Genome 8:866-868(1997).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
CC SYSTEM.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Y13436; CAA73847.1; -;
CC DR HSSP; Q05066; IHRZ.
CC DR MIM; 602148; -;
CC DR InterPro; IPR000910; -;
CC DR Pfam; PF00505; HMG_box; 1.
CC KW DNA-binding; Nuclear protein.
FT DOMAIN 27 43 POLY-GLY.
FT DNA_BIND 51 119 HMG BOX.
FT DOMAIN 145 150 POLY-GLY.
FT DOMAIN 197 204 POLY-ALA.
FT DOMAIN 280 287 POLY-ALA.
FT DOMAIN 292 302 POLY-ALA.
FT DOMAIN 353 360 POLY-ALA.
SQ SEQUENCE 387 AA; 38855 MW; 8D43A38002CE3494 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 387;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 AVAAAAAA 20
Db 292 AVAAAAAA 299
|||||||
RESULT 13
HME1_HUMAN STANDARD; PRT; 391 AA.
ID HME1_HUMAN AC Q05925;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HOMEBOX PROTEIN ENGRAILED-1 (HU-EN-1).
GN EN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93185339; PubMed=1363401;
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
RA Provart N.J., Joyner A.L.;
RT "Cloning and sequence comparison of the mouse, human, and chicken
RT engrailed genes reveal potential functional domains and regulatory
RT regions.";

RL Dev. Genet. 13:345-358(1992).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- SIMILARITY: BELONGS TO THE ENGRAILED FAMILY OF HOMEOBOX PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L12698; AAA53501.1; -;
 CC EMBL; L12699; AAA53502.1; -;
 CC HSSP; P02836; 1ENH.
 CC TRANSFAC; T02015; -;
 CC MTM; L31290; -;
 CC InterPro; IPR000747; -;
 CC InterPro; IPR001356; -;
 CC Pfam; PF00046; homeobox; 1.
 CC PRINTS; PR00024; HOMEOBOX.
 CC PRINTS; PR00026; ENGRAILED.
 CC PROSITE; PS00027; HOMEOBOX_1; 1.
 CC PROSITE; PS00071; HOMEOBOX_2; 1.
 CC PROSITE; PS00033; ENGRAILED; 1.
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 KW Homeobox; DNA-binding; ENGRAILED; 1.
 FT DOMAIN 199 218 POLY-ALA.
 FT DOMAIN 224 231 POLY-GLY.
 FT DNA_BIND 302 361 HOMEOBOX.
 FT SEQUENCE 391 AA; 40044 MW; 30DA29BF0995956F CRC64;
 SQ
 Query Match 2.3%; Score 8; DB 1; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 AVAAAAAA 20
 DB 208 AVAAAAAA 215
 RESULT 14
 SOX1_MOUSE STANDARD; PRT; 391 AA.
 ID SOX1_MOUSE
 AC P53783;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SOX-1 PROTEIN.
 GN SOX1 OR SOX-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=96189340; PubMed=8625802;
 RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,
 RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,
 RA Lovell-Badge R.;
 RT "A comparison of the properties of Sox-3 with Sry and two related
 RT genes, Sox-1 and Sox-2";
 RL Development 122:509-520(1996).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -!- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
 CC SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.
 CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X94126; CAA63846.1; -;
 CC HSSP; Q05066; LHR2.
 CC MGD; MGI:98357; Sox1.
 CC InterPro; IPR000910; -;
 CC Pfam; PF00505; HMG-box; 1.
 CC DNA-binding; Nuclear protein.
 KW DNA-binding; Nuclear protein.
 FT DOMAIN 30 43 POLY-GLY.
 FT DNA_BIND 51 119 HMG BOX.
 FT DOMAIN 145 150 POLY-GLY.
 FT DOMAIN 197 204 POLY-ALA.
 FT DOMAIN 280 288 POLY-ALA.
 FT DOMAIN 296 306 POLY-ALA.
 FT DOMAIN 357 364 POLY-ALA.
 FT SEQUENCE 391 AA; 39237 MW; 9F81ED667F947C05 CRC64;
 SQ
 Query Match 2.3%; Score 8; DB 1; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 AVAAAAAA 20
 DB 296 AVAAAAAA 303
 RESULT 15
 HMEL_MOUSE STANDARD; PRT; 401 AA.
 ID HMEL_MOUSE
 AC P09065;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HOMEOBOX PROTEIN ENGRAILED-1 (MO-EN-1).
 GN EN1 OR EN-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93185339; PubMed=1363401;
 RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
 RA Provart N.J., Joyner A.L.;
 RT "Cloning and sequence comparison of the mouse, human, and chicken
 RT engrailed genes reveal potential functional domains and regulatory
 RT regions.";
 RL Dev. Genet. 13:345-358(1992).
 RN [2]
 RP SEQUENCE OF 278-401 FROM N.A.
 RX MEDLINE=88112776; PubMed=2892757;
 RA Joyner A.L., Martin G.R.;
 RT "En-1 and En-2, two mouse genes with sequence homology to the
 RT Drosophila engrailed gene: expression during embryogenesis.";
 RL Genes Dev. 1:29-38(1987).
 RN [3]
 RP SEQUENCE OF 298-401 FROM N.A.
 RX MEDLINE=86079501; PubMed=2416459;
 RA Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;
 RT "Expression during embryogenesis of a mouse gene with sequence
 RT homology to the Drosophila engrailed gene.";
 RL Cell 43:29-37(1985).
 RN [4]
 RP SEQUENCE OF 321-380 FROM N.A.
 RX MEDLINE=91099509; PubMed=1980115;
 RA Holland P.W.H., Williams N.A.;
 RT "Conservation of engrailed-like homeobox sequences during vertebrate
 RT evolution.";
 RL FEBS Lett. 277:250-252(1990).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.

```

CC -1- SIMILARITY: BELONGS TO THE ENGRAILED FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L12703; AAA03660.2; -
CC EMBL; Y00201; CAA68361.1; -
CC PIR; A26629; A26629.
CC PIR; A24778; A24778.
CC PIR; S13009; S13009.
CC PIR; A48423; A48423.
CC TRANSFAC; T02016; -
CC MGD; MGI:95389; Enl.
CC InterPro; IPR000747; -
CC InterPro; IPR001356; -
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC PRINTS; PR00026; ENGRAILED.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC PROSITE; PS00033; ENGRAILED; 1.
CC HOMEBOX; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 52 87 PRO-RICH.
FT DOMAIN 73 87 POLY-PRO.
FT DOMAIN 207 228 POLY-ALA.
FT DNA_BIND 312 371 HOMEBOX.
SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152FAE CRC64;

```

```

Query Match      2.3%; Score 8; DB 1; Length 401;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 13 AVAAAAAA 20
   |||||
Db 220 AVAAAAAA 227

```

Search completed: August 23, 2001, 11:18:30
Job time: 197 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:19:26 ; Search time 50.81 Seconds
(without alignments)
908.767 Million cell updates/sec

Title: US-09-654-652A-3

Perfect score: 349

Sequence: 1 MNIKTAVKSALAVAAAAA.....AKGAKVNPNGHKRYRVNFHEH 349

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: SPTRMBL_16.*

2: sp_archaea.*

3: sp_bacteria.*

4: sp_fungi.*

5: sp_human.*

6: sp_invertebrate.*

7: sp_mmc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_unclassified.*

13: sp_vertebrate.*

14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	3.2	539	10 Q9SR71	Q9sr71 arabidopsis
2	10	2.9	250	10 Q9LJ08	Q9lj08 oryza sativ
3	10	2.9	348	10 Q96517	Q96517 arabidopsis
4	10	2.9	348	10 Q96517	Q96517 arabidopsis
5	10	2.9	369	10 Q96517	Q96517 arabidopsis
6	10	2.9	621	2 Q93301	Q93301 cellvibrio
7	9	2.6	87	10 Q9LJ13	Q9lj13 oryza sativ
8	9	2.6	184	1 Q9HQS5	Q9hqs5 halobacteri
9	9	2.6	300	2 Q9KWF8	Q9kwf8 kitasatospo
10	9	2.6	321	7 Q9XS19	Q9xs19 papio hamad
11	9	2.6	350	2 Q9KMG4	Q9kmg4 streptomyce
12	9	2.6	439	1 Q9YCK0	Q9yck0 aeropyrum p
13	9	2.6	518	10 Q9PRC3	Q9prc3 oryza sativ
14	9	2.6	776	2 Q86516	Q86516 streptomyce
15	9	2.6	840	5 Q9NB03	Q9nb03 aedes aegyp
16	9	2.6	1068	5 Q9W3X2	Q9w3x2 drosophila
17	9	2.6	1307	5 Q9V702	Q9v702 drosophila
18	9	2.6	3112	5 Q9NKP1	Q9nkp1 leishmania
19	8	2.3	23	4 Q43519	Q43519 homo sapien

20 8 2.3 38 10 Q9S9D9
21 8 2.3 46 7 P79534
22 8 2.3 46 7 P79536
23 8 2.3 46 7 P79538
24 8 2.3 46 7 P79538
25 8 2.3 67 5 Q9NNB6
26 8 2.3 82 13 Q9YH51
27 8 2.3 82 13 Q99013
28 8 2.3 101 10 Q9LGG6
29 8 2.3 107 4 Q16846
30 8 2.3 111 4 Q15587
31 8 2.3 134 4 Q43816
32 8 2.3 135 4 Q9NP14
33 8 2.3 137 6 Q29365
34 8 2.3 138 4 P78428
35 8 2.3 139 6 Q9GLV8
36 8 2.3 141 7 Q9GJ06
37 8 2.3 152 4 Q00496
38 8 2.3 152 5 Q9U730
39 8 2.3 152 5 Q9TVX1
40 8 2.3 153 5 Q9U731
41 8 2.3 153 5 Q9TW37
42 8 2.3 154 5 Q9TW44
43 8 2.3 158 5 Q02055
44 8 2.3 175 10 Q9LGB7
45 8 2.3 197 2 Q9X6J7
198 10 Q9FTW8

Q9S9d9 nicotiana t
P79534 homo sapien
P79536 homo sapien
P79538 homo sapien
Q9nnb6 leishmania
Q9yh51 pseudopleur
Q99013 pseudopleur
Q9lgg6 oryza sativ
Q16846 homo sapien
Q15587 homo sapien
Q43816 homo sapien
Q9np14 homo sapien
Q29365 sus scrofa
P78428 homo sapien
Q9glv8 equus cabal
Q9gj06 homo sapien
Q00496 homo sapien
Q9u730 drosophila
Q9tvx1 drosophila
Q9u731 drosophila
Q9tw37 drosophila
Q9tw44 drosophila
Q02055 caenorhabdi
Q9lgb7 oryza sativ
Q9x6j7 vibrio chol
Q9ftw8 oryza sativ

ALIGNMENTS

RESULT 1

Q9SR71 ID Q9SR71 PRELIMINARY; PRT; 539 AA.
AC Q9SR71;
DT 01-MAY-2000 (T22K18.10, Created)
DT 01-MAY-2000 (T22K18.10, Last sequence update)
DT 01-JUN-2000 (T22K18.10, Last annotation update)
DE T22K18.10 PROTEIN.
GN T22K18.10.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010927; AAF04417.1;
DR InterPro; IPR000166;
DR InterPro; IPR002965;
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 539 AA; 57710 MW; 72597A2E1B57EE9B CRC64;

Query Match 3.2%; Score 11; DB 10; Length 539;
Best Local Similarity 100.0%; Pred. No. 0.049;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SSSSAPASSSS 283

| | | | | | | | | |

Db 117 SSSSAPASSSS 127

RESULT 2

Q9LJ08 ID Q9LJ08 PRELIMINARY; PRT; 250 AA.

AC Q9LJ08;

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE ESTS AU067519(C10906).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 OC Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0667A10.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF001073; BAA89582.1; -;
 DR InterPro; IPR000010; -;
 DR Pfam; PF00031; cystatin; 2.
 DR PROSITE; PS00287; CYSTATIN; UNKNOWN_1.
 DR SMART; SM00043; CY; 1.
 SQ SEQUENCE 250 AA; 27269 MW; 6F2A7D4EB8AF66F CRC64;

Query Match 2.9%; Score 10; DB 10; Length 250;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAAL 21
 DB 24 LAVAAAAAL 33

RESULT 3
 Q96517 ID Q96517 PRELIMINARY; PRT; 348 AA.
 AC Q96517;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE MAP KINASE KINASE ALPHA PROTEIN KINASE (EC 2.7.1.37) (PHOSPHORYLASE B
 DE KINASE KINASE) (GLYCOGEN SYNTHASE A KINASE) (HYDROXYALKYL-PROTEIN
 DE KINASE) (SERINE(THREONINE) PROTEIN KINASE).
 GN MAP2KALPHA.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97092864; PubMed=8938426;
 RA Jouannic S., Hamal A., Kreis M., Henry Y.;
 RL Plant Physiol. 112:1397-1397(1996).
 DR EMBL; Y07694; CAA68958.1; -;
 DR Mendel; 14316; Arath:2349;14316.
 DR InterPro; IPR000719; -;
 DR InterPro; IPR002290; -;
 DR Pfam; PF00069; pkinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR SMART; SM00220; S_TKc; 1.
 KW ATP-binding; Kinase; Transferase.
 SQ SEQUENCE 348 AA; 38329 MW; C18C987C6B7AF5F5 CRC64;

Query Match 2.9%; Score 10; DB 10; Length 348;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SSSAPASSS 282
 DB 46 SSSAPASSS 55

RESULT 4
 O80398 ID O80398 PRELIMINARY; PRT; 348 AA.
 AC O80398;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE MAP KINASE KINASE 5.
 GN MKK5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Shinozaki K., Ichimura K., Mizoguchi T.;
 RT "Molecular cloning and characterization of cDNAs encoding mitogen-
 RT activated protein kinase kinase in Arabidopsis thaliana.";
 RL submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 CC -I- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AB015316; BAA28831.1; -;
 DR EMBL; AB023045; BAB01714.1; -;
 DR HSP; P24941; ICKP.
 DR Mendel; 31061; Arath:2349;31061.
 DR InterPro; IPR000719; -;
 DR InterPro; IPR002290; -;
 DR Pfam; PF00069; pkinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR SMART; SM00220; S_TKc; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 348 AA; 38329 MW; 216C96728F34B5B5 CRC64;

Query Match 2.9%; Score 10; DB 10; Length 348;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SSSAPASSS 282
 DB 46 SSSAPASSS 55

RESULT 5
 O04322 ID O04322 PRELIMINARY; PRT; 369 AA.
 AC O04322;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE MYB-RELATED TRANSCRIPTION ACTIVATOR (MYBST1) ISOLOG (MYBST1).
 GN T02004.16.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OX Brassicaceae; Arabidopsids.
[1]
RN NCBI_TaxID=3702;
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Phillips C.A., Brandon R.C.,
RA Fuhrmann J.L., White O., Kerlavage A.R., Adams M.D., Somerville C.R.,
RL Venter J.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AC001645; AB63650.1; -
DR EMBL; AB023046; BAB01274.1; -
DR Mendel; 17078; Arath;2095:17078.
DR InterPro; IPR001005; -
DR Pfam; PF00249; mvb_DNA-binding; 1.
DR PROSITE; PS0090; MYB.3.1.
SQ SEQUENCE 369 AA; 40043 MW; DCC039507044AC9E CRC64;

Query Match 2.9%; Score 10; DB 10; Length 369;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAAA 19
| | | | | | | |
Db 52 SALAVAAAAA 61

RESULT 6
Q59301
ID Q59301 PRELIMINARY; PRT; 621 AA.
AC Q59301;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ENDO-BETA-1,4-XYLANASE PRECURSOR (EC 3.2.1.8) (ENDO-1,4-BETA-XYLANASE)
DE (1,4-BETA-D-XYLAN XYLANOXYDROLASE).
OS Cellvibrio mixtus.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Cellvibrio.
OX NCBI_TaxID=39650;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MIXTUS;
RX MEDLINE=96077124; PubMed=7492333;
RA Millward-Sadler S.J., Davidson K., Hazlewood G.P., Black G.W.,
RA Gilbert H.J., Clarke J.H.;
RT "Novel cellulose-binding domains, NodB homologues and conserved
RT modular architecture in xylanases from the aerobic soil bacteria
RT Pseudomonas fluorescens subsp. cellulosa and Cellvibrio mixtus.";
RL Biochem. J. 312:39-48(1995).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC LINKAGES IN XYLANS.
DR EMBL; Z48926; CAA88762.1; -
DR HSSP; P14768; ICLX.
DR InterPro; IPR001000; -
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; UNKNOWN_1.

KW Signal; Xylan degradation; Hydrolase; Glycosidase.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 621
SQ SEQUENCE 621 AA; 64929 MW; 830E5B959D2C3AB8 CRC64;

Query Match 2.9%; Score 10; DB 2; Length 621;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 SSSAPASSSS 283
| | | | | | | |
Db 89 SSSAPASSSS 98

RESULT 7
Q9LL13
ID Q9LL13 PRELIMINARY; PRT; 87 AA.
AC Q9LL13;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN (OSJNBA0086P08.3 PROTEIN).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0708G02.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:OSJNBA0086P08.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001539; BAA92926.1; -
DR EMBL; AP002855; BAB17194.1; -
SQ SEQUENCE 87 AA; 9123 MW; F3ACD82BA63824EE CRC64;

Query Match 2.6%; Score 9; DB 10; Length 87;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAA 20
| | | | | | | |
Db 22 LAVAAAAA 30

RESULT 8
Q9HQS5
ID Q9HQS5 PRELIMINARY; PRT; 184 AA.
AC Q9HQS5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE VNG1026H.
GN VNG1026H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,

RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Daie H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL: AE005037; AG19438.1; -;
 SQ SEQUENCE 184 AA; 18956 MW; B120BC2158DD9095 CRC64;

Query Match 2.6%; Score 9; DB 1; Length 184;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAA 20
 |||||
 Db 137 LAVAAAAA 145

RESULT 9
 Q9KWF8 ID Q9KWF8 PRELIMINARY; PRT; 300 AA.
 AC Q9KWF8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE MEVALONATE DIPHOSPHATE DECARBOXYLASE.
 GN MDPO.
 OS Kitasatospora griseola.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Kitasatospora.
 OX NCBI_TaxID=2064;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Daiji T.;
 RT "Cloning of a gene cluster encoding enzymes responsible for the
 RT mevalonate pathway from a terpenoid antibiotic-producing Streptomyces
 RT strain";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB037907; BAB07818.1; -;
 DR InterPro: IPR001745; -;
 DR Pfam: PF00288; GHMP_Kinases; 1.
 SQ SEQUENCE 300 AA; 30865 MW; 17D9F9D707781997 CRC64;

Query Match 2.6%; Score 9; DB 2; Length 300;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ALAVAAAA 19
 |||||
 Db 78 ALAVAAAA 86

RESULT 10
 Q9XS19 ID Q9XS19 PRELIMINARY; PRT; 321 AA.
 AC Q9XS19;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE MHC CLASS I RELATED PROTEIN PRECURSOR (FRAGMENT).
 GN MIC1.
 OS Papio hamadryas (Hamadryas baboon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Papio.
 OX NCBI_TaxID=9557;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Seo J.W., Ross C., Bontrop R., Hunsmann G., Walter L., Guenther E.;
 RT "Major histocompatibility complex-linked MIC genes in rhesus macaques
 RT and other primates";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
 CC IMMUNE SYSTEM (BY SIMILARITY).
 CC -I- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
 CC MICROGLOBULIN) (BY SIMILARITY).
 CC -I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 DR EMBL: AJ242442; CAB46518.1; -;
 DR InterPro: IPR001039; -;
 DR InterPro: IPR003006; -;
 DR InterPro: IPR003597; -;
 DR Pfam: PF00047; Ig; 1.
 DR Pfam: PF00129; MHC_I; 1.
 DR PRODOM: PD000050; -; 1.
 DR SMART: SM00407; IgC1; 1.
 KW Glycoprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 321
 SQ SEQUENCE 321 AA; 36039 MW; 8B2579752CE6B65F CRC64;

Query Match 2.6%; Score 9; DB 7; Length 321;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAA 20
 |||||
 Db 291 LAVAAAAA 299

RESULT 11
 Q9KWG4 ID Q9KWG4 PRELIMINARY; PRT; 350 AA.
 AC Q9KWG4;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE DIPHOSPHOMEVALONATE DECARBOXYLASE.
 OS Streptomyces sp. (strain CL190).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=93372;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CL190;
 RX MEDLINE=20353446; PubMed=10894721;
 RA Takagi M., Kuzuyama T., Takahashi S., Seto H.;
 RT "A gene cluster for the mevalonate pathway from Streptomyces sp.
 RT strain CL190";
 RL J. Bacteriol. 182:4153-4157(2000).
 DR EMBL: AB037666; BAB07791.1; -;
 DR InterPro: IPR001459; -;
 DR InterPro: IPR001745; -;
 DR Pfam: PF00288; GHMP_Kinases; 1.
 DR PRINTS: PR00959; MEVGALKINASE.
 SQ SEQUENCE 350 AA; 36409 MW; 1CB61F306ACDC7CF CRC64;

Query Match 2.6%; Score 9; DB 2; Length 350;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ALAVAAAA 19
 |||||
 Db 130 ALAVAAAA 138

RESULT 12
 Q9YCK0 ID Q9YCK0 PRELIMINARY; PRT; 439 AA.

AC O9YCK0;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 45.2 KDA PROTEIN APE1257.
 GN APE1257.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000061; BAA80247.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 439 AA; 45223 MW; 935D70CED06CF1EA CRC64;
 Query Match 2.6%; Score 9; DB 1; Length 439;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 SALAVAAAA 18
 | | | | | | | |
 DB 398 SALAVAAAA 406
 RESULT 13
 Q9FRC3 PRELIMINARY; PRT; 518 AA.
 AC Q9FRC3;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE PUTATIVE CYTOCHROME P450-RELATED PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 OC Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
 RA Zismann V., Pal G., Bowman C.L., Fujii C.Y., VanAken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 3 BAC OSJNBa0013M12 genomic sequence.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC082644; AAC46147.1; -;
 SQ SEQUENCE 518 AA; 58749 MW; E3B82EDD23183C81 CRC64;

Query Match 2.6%; Score 9; DB 10; Length 518;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 LAVAAAAA 20
 | | | | | | | |
 DB 7 LAVAAAAA 15
 RESULT 14

O86516 PRELIMINARY; PRT; 776 AA.
 AC O86516;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DE PUTATIVE SECRETED LYASE.
 GN SCIC2.15.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL031124; CAA19982.1; -;
 DR InterPro: IPR003159; -;
 DR Pfam: PF02278; Lyase_8; 1.
 KW Lyase.
 FT DOMAIN 17 23 POLY-ALA.
 FT DOMAIN 30 35 POLY-ALA.
 SQ SEQUENCE 776 AA; 83774 MW; DE861687595B29E1 CRC64;
 Query Match 2.6%; Score 9; DB 2; Length 776;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 AAAAAALTT 23
 | | | | | | | |
 DB 18 AAAAAALTT 26
 RESULT 15
 Q9NB03 PRELIMINARY; PRT; 840 AA.
 AC Q9NB03;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE NUCLEAR RECEPTOR FTZ-F1 PROTEIN.
 GN FTZ-F1.
 OS Aedes aegypti (Yellowfever mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
 OC Aedes.
 OX NCBI_TaxID=7159;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Li C., Kapitskaya M.Z., Zhu J., Miura K., Segraves W., Raikhel A.S.;
 RT "Conserved Molecular Mechanism for the Stage Specificity of the
 RT Mosquito Vitellogenin Response to Ecdysone.";
 RL Dev. Biol. 0:0-0(2000).
 CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -|- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
 DR EMBL: AF274870; AAF82307.1; -;
 DR InterPro: IPR000536; -;

DR InterPro: IPR001628; -.
 DR InterPro: IPR001723; -.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00105; zf-C4; 1.
 DR PRINTS: PR00398; STRDHORMONER.
 DR PRINTS: PR00047; STROIDFINGER.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 DR SMART: SM00430; HOLI; 1.
 KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
 KW Zinc-finger.
 SQ SEQUENCE 840 AA; 89623 MW; AC85DB8DCE2DDA3A CRC64;

Query Match 2.6%; Score 9; DB 5; Length 840;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 AVAAAAAAL 21
 |||||
 Db 191 AVAAAAAAL 199

Search completed: August 23, 2001, 11:19:27
 Job time: 199 sec


```

|||||
Db 937 AGAGATGACGAACCTGCTCCG 957

RESULT 2
LOCUS BSPGALG 1416 bp DNA BCT 07-NOV-1994
DEFINITION Bacillus sp. bgaA gene for lichenase.
ACCESSION Z12151
VERSION 212151.1 GI:296931
KEYWORDS bgaI gene; lichenase.
SOURCE Bacillus sp.
ORGANISM Bacillus sp.
REFERENCE 1 (bases 1 to 1416)
AUTHORS Tabernero,C., Coll,P.M., Fernandez-Abalos,J.M., Perez,P. and
          Santamaria,R.I.
TITLE Cloning and DNA sequencing of bgaA, a gene encoding an
       endo-beta-1,3-1,4-glucanase, from an alkalophilic Bacillus strain
       (N137)
JOURNAL Appl. Environ. Microbiol. 60 (4), 1213-1220 (1994)
MEDLINE 94288605
REFERENCE 2 (bases 1 to 1416)
AUTHORS Santamaria,R.I.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1992) Ramon I. Santamaria, Instituto de
       Microbiologia Bioquimica, C.S.I.C./ Universidad, de Salamanca,
       Plaza de la Merced S/N, Salamanca, Salamanca, 37008, Spain
FEATURES
       source
           1..1416
               /organism="Bacillus sp."
               /strain="137"
               /db_xref="taxon:1409"
               /clone="10"
               199..204
               -35_signal
               -10_signal
               5'UTR
               RBS
               gene
               CDS
               199..204
               /transl_table=11
               /product="lichenase"
               /protein_id="CAA78135.1"
               /db_xref="GI:296932"
               /db_xref="SPTREMBL:Q45648"
               /translation="MKKLVLFSCMLLLFSLISGLVSPQVAEAAERPIGTAFTVET
               ESYDSRWKAGVWSNGQMFNATWYPEQVTSNGKMKLOIEKEDNETASPPYKAGELR
               TNDVHYGLFEYSMKPAKSTGTVSFFTYTGPWDNDPDEIDIEFLGKDTKIOFN
               YFTNGVGGNEHYHELGFDAADENTYAFERPESTPWFVNGELVHTATENIQTPQKI
               MNNLPGIGVDGTRGFRNGEDIPVTQYDWWKYTPLEELGCTNEKNNKTKCKTKVK
               "
               388..480
               sig_peptide
               481..1215
               mat_peptide
               /gene="bgaI"
               /gene="bgaI"
               /product="lichenase"
               509 a 181 c 276 g 450 t

Query Match 6.7%; Score 50; DB 2; Length 1416;
Best Local Similarity 51.5%; Pred. No. 0.0055;
Matches 172; Conservative 0; Mismatches 150; Indels 12; Gaps 2;

Qy 25 ggtgcgaactctacacgttagaagaagttcagtcacgttgattgaagcccgatgaag 84
|||||
Db 679 GCTGAGAACTCCACAAACGATTTTATCATACGGGTGTTTGAAGTCAGTATGAAA 738
|||||
Qy 85 atggcagccgcatcggggaacagtcagttccatgttcctaccagaatggttcggaatc 144
|||||
Db 739 CTTGCAAAAGTCAACGGGAACAGTCTCTTCATTTTACCTATACTGACCTTGGATTGG 798
|||||
Qy 145 gccgatggaagggccctgggtagaagtggatattgaagtctctcggaagaatccgggcagt 204
|||||
Db 799 GAGAAATGAT--CCATGGGATGAATGATATCGAATTTTAGTAAAGTACTACTATAA 855
|||||
Qy 205 ttccagtcacaatcattaccggttaagccggcggaacaaagactagcgaagaagcaccat 264
|||||
Db 856 ATACAATTTAATTATTTACAAACGGAGTAGGGCGGAATGAGCATT-----ACCAT 906
|||||
Qy 265 gctgttagcccccgcgcgcaggtcttcacacacctacggtctcgaatgactccgaat 324
|||||
Db 907 GAATTAGGATTGATGACGAGATGATTTTAATACGATGCTTTTGTAGTGAGACCAGAA 966
|||||
Qy 325 taagtccgctggactgttgacggtcaggaaagtc 358
|||||
Db 967 TCTATTGCTTGGTTGTAATGTTGAATTGGTTC 1000
|||||

SYHA78M 1234 bp DNA SYN 04-APR-2000
LOCUS Synthetic B.macerans/B.amyloliquefaciens hybrid gene for
DEFINITION endo-1,3-1,4-beta-glucanase.
ACCESSION Z25882
VERSION 225882.1 GI:398202
KEYWORDS beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 777)
AUTHORS Hofemeister,J., Kurtz,A., Borriss,R. and Knowles,J.
TITLE The beta-glucanase gene from Bacillus amyloliquefaciens shows
       extensive homology with that of Bacillus subtilis
JOURNAL Gene 49 (2), 177-187 (1986)
MEDLINE 87192007
REFERENCE 2 (bases 396 to 852)
AUTHORS Borriss,R., Buettner,K. and Maentsaelae,P.
TITLE Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
       homologues to other beta-glucanases
JOURNAL Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
MEDLINE 91109712
REFERENCE 3 (bases 1 to 1234)
AUTHORS Olsen,O., Borriss,R., Simon,O. and Thomsen,K.K.
TITLE Hybrid Bacillus (1-3-1,4)-beta-glucanases: engineering thermostable
       enzymes by construction of hybrid genes
JOURNAL Mol. Gen. Genet. 225 (2), 177-185 (1991)
MEDLINE 91172113
REFERENCE 4 (bases 1 to 1234)
AUTHORS Pollitz,O., Simon,O., Olsen,O. and Borriss,R.
TITLE Determinants for the enhanced thermostability of hybrid
       (1-3,1-4)-beta-glucanases
JOURNAL Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE 94009045
REFERENCE 5 (bases 1 to 1234)
AUTHORS Pollitz,O.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1993) Pollitz O., Humboldt University of Berlin,
       Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
FEATURES
       source
           1..1234
               /organism="synthetic construct"
               /db_xref="taxon:32630"
               /focus
               1..1234
               /organism="Bacillus amyloliquefaciens"
               /db_xref="taxon:1390"
               1..777
               /organism="Bacillus amyloliquefaciens"
               /db_xref="taxon:1390"
               1..1234
               /source
       source
           1..1234
               /source
       source
           1..1234
               /source

```


JOURNAL Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin,
Biology, Warschauer Str. 43, BERLIN, Germany, D-10243

FEATURES
source Location/Qualifiers
1..1228
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /focus
source 1..1228
 /organism="Bacillus amyloliquefaciens"
 /db_xref="taxon:1390"
source 1..543
 /organism="Bacillus amyloliquefaciens"
 /db_xref="taxon:1390"
source 1..1228
 /organism="Paenibacillus macerans"
 /db_xref="taxon:44252"
sig_peptide 469..543
CDS 469..1182
 /function="endo-1,3-1,4-beta-glucanase"
 /note="Hybrid between bgla (Pos. 1 to 543) and bglm
 (Pos.in hybrid 544 to 1228, Pos. Original 168 to 852),"
 /codon_start=1
 /transl_table=11
 /label=HAOM
 /evidence=experimental
 /product="hybrid-endo-1,3-1,4-beta-glucanase"
 /protein_id="CAA81092.1"
 /db_xref="GI:398186"
translation="MKRVLLILVTGLFMSLCGTTSSVSAGSVFWELPSVFNPNSTWPKA
DYSNGGVNCTWRANNVNFTNDGKLGLTISAYNKFAEYRSTNIYGYGLEYFSVM
FKPAGTIVSYSFYTGPAHQWDIEFLGKDTTFQFYNYTVNGVGSGHEKVLSL
FDASGKHFFATDWQPGYIKWYVDGVKLKTATANIPISTFGKIMMNLWGNTGVDWLIGS
YGNAPLYAEYDWVKYTSN"
544..1228
 /organism="Paenibacillus macerans"
 /db_xref="taxon:44252"
BASE COUNT 371 a 220 c 290 g 347 t
ORIGIN

Query Match 6.0%; Score 44.6; DB 56; Length 1228;
Best Local Similarity 52.9%; Pred. No. 0.17;
Matches 154; Conservative 0; Mismatches 119; Indels 18; Gaps 2;

Qy 58 tacggttaagttagcccgctgaagatggcaggccgatcgccgcatacggaaacagtcagttccatg 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 TAGGGCGCTGTACGAGGTCACTATGAACGCACGCCAAAATAACAGGAATGTCTCATCCTTT 810
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 118 ttocetaccagaatggttccgaaaatgccgatggaggccctgggttagaagtggatatatt 177
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 TTCACGTA-----TACAGGACCTGTCTCATGGCACAAATGGGATGAATAGATATC 861
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 178 gaagttctcgaagaatccggcgagtttccagtcacaatcatattaccggttaagccgcg 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 862 GAATTTCAGGAAAGACACGCAAAAGTCCAGTTTAACATTATATACCAATGGGGTTGGC 921
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 238 gcacaaaagactagcaaaaaacaccatgctgttgagccccgcgcgatcaggtcttccac 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 922 GGTCAT-----GAAAGGTTATCTCTCTGGCTTTTATGCATCAAGAGGGCTTCAT 972
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 298 acctacggtctcgaatggactccgaattaacgtccgctggagctgttacggt 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 973 ACCTATGCTTTGCTATGGCAGGCAGGGGTATATTAATGGTATGTAGACGCT 1023
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
SYHA12Y13
LOCUS SYHA12Y13 1234 bp DNA SYN 04-APR-2000
DEFINITION Synthetic B.macerans/B. amyloliquefaciens hybrid gene for
 endo-1,3-1,4-beta-glucanase.
ACCESSION Z25876
VERSION Z25876.1 GI:398190
KEYWORDS beta-glucanase; endo-1,3-1,4-beta-glucanase.

SOURCE	ORGANISM	synthetic construct.
REFERENCE	AUTHORS	synthetic construct
REFERENCE	TITLE	artificial sequence.
JOURNAL	MEDLINE	1 (bases 1 to 579)
REFERENCE	AUTHORS	Hofemeister,J., Kurtz,A., Borriass,R. and Knowles,J.
REFERENCE	TITLE	The beta-glucanase gene from <i>Bacillus amyloliquefaciens</i> shows extensive homology with that of <i>Bacillus subtilis</i> Gene 49 (2), 177-187 (1986)
JOURNAL	MEDLINE	87192007
REFERENCE	AUTHORS	2 (bases 198 to 852)
REFERENCE	TITLE	Borriass,R., Buettner,K. and Maentsaelae,P.
JOURNAL	MEDLINE	Structure of the beta-1,3-1,4-glucanase gene of <i>Bacillus macerans</i> : (1-3,1-4)-beta-glucanases
REFERENCE	AUTHORS	Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
REFERENCE	TITLE	91109712
JOURNAL	MEDLINE	3 (bases 1 to 1234)
REFERENCE	AUTHORS	Politz,O., Simon,O., Olsen,O. and Borriass,R.
REFERENCE	TITLE	Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases
JOURNAL	MEDLINE	Eur. J. Biochem. 216 (3), 829-834 (1993)
REFERENCE	AUTHORS	4 (bases 1 to 1234)
REFERENCE	TITLE	Politz,O., Simon,O., Olsen,O. and Borriass,R.
JOURNAL	MEDLINE	Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases
REFERENCE	AUTHORS	Eur. J. Biochem. (1993) In press
REFERENCE	TITLE	5 (bases 1 to 1234)
JOURNAL	MEDLINE	Politz,O.
REFERENCE	AUTHORS	Direct Submission
JOURNAL	MEDLINE	Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
FEATURES	Source	Location/Qualifiers
		1. .1234
		/organism="synthetic construct"
		/db_xref="taxon:32630"
		/focus
	source	1. .1234
		/organism="Bacillus amyloliquefaciens"
		/db_xref="taxon:1390"
	source	1. .579
		/organism="Bacillus amyloliquefaciens"
		/db_xref="taxon:1390"
	source	1. .1234
		/organism="Paenibacillus macerans"
		/db_xref="taxon:44252"
	sig_peptide	469..543
	CDS	469..1188
		/function="endo-1,3-1,4-beta-glucanase"
		/notes="Hybrid between bgla (Pos. 1 to 579) and bglm (Pos. in hybrid 580 to 1234, Pos. original 198 to 852) Deletion of Y13 of HA12-M"
		/codon_start=1
		/transl_table=1
		/label=HA12M
		/evidence=experimental
		/product="hybrid-endo-1,3-1,4-beta-glucanase"
		/protein_id="CAA81095.1"
		/db_xref="GI:398191"
		/translation="MKRVLLILVTGLFMSLCGTSITSSVAQTGGSPFPENSPNPSTW KADGYNGGVNFCWNRANNVNTDGLKLTSSAYNKFDCAEYRTNIYGYLVEY SMKPAKNTGIVSSFFTYTPGHCOTWDEIDIEFLGKDTTKVQFNYYTNGVGGHKEVLS LGFDASKGHPTYAFDQPGYIKWYVDGLKHTATANIPTPGKIMNLNNGTGVDDDL GSYNGANPLYAEYDWNKYTSN"
	source	580..1234
		/organism="Paenibacillus macerans"
		/db_xref="taxon:44252"
BASE COUNT	374 a	225 c 287 g 348 t
ORIGIN		

Query Match 6.0%; Score 44.6; DB 56; Length 1234;
Best Local Similarity 52.9%; Pred. No. 0.17;

[illegible]

1

```

source      1. 852
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
93...806
/codon_start=1
/transl_table=11
/product="1,3-1,4-glucanase"
/protein_id="CAA39426.1"
/db_xref="GI:296716"
/db_xref="SWISS-PROT:P23904"
/translacion="MKKSCFTLVTFAPLSFVSALAGSVFWEPLSVFNSTWEKA
DYGNGVNCFTWRNVNFTNDGKLGLTSSAYNKFPDCAEYRSTNIYGYGLYEVSM
KPAKNTGIVSSFTYTGPAHGTQWDEIDIEFLGKDTTKVQFNYYNGVGHKVISLG
FPAKGFHYAFDMQPGYIKWYVDGVLKHTATANIPSTPGKIMNMLNNGTGVDDWUGS
YNGANPLYAEYDMVKYTSN"
BASE COUNT 248 a 156 c 203 g 245 t
ORIGIN

Query Match      5.8%; Score 43; DB 2; Length 852;
Best Local Similarity 52.6%; Pred. No. 0.46;
Matches 153; Conservative 0; Mismatches 120; Indels 18; Gaps 2;

QY 58 tacgtaagttaagccgqatgaagatggcagccgcatcggaacagtcagttccatg 117
||||| || || || || || || || || || || || || || || || || || ||
Db 375 TACGGCGCTGACAGGTCAGTATGAAGCCAGCCAAAAATACAGGAATGTGTCATCCTTT 434
||||| || || || || || || || || || || || || || || || || || ||

QY 118 ttctctaccagaatggttcggaatcgccgcatggaagccctgggtagaagtggatatt 177
||||| || || || || || || || || || || || || || || || || || ||
Db 435 TTACAGTA-----TACAGACCTGTCATCGGCACACATGGATGGAATAGATATC 485
||||| || || || || || || || || || || || || || || || || || ||

QY 178 gaagttctcggaagaatccgggcagtttcagtcaccaatcattaccggtaagccgcgc 237
||||| || || || || || || || || || || || || || || || || || ||
Db 486 GAATTTCTAGAAAGACACAGCAAAAGTCAGCTTTAACTATTATACCAATGGGTTGGC 545
||||| || || || || || || || || || || || || || || || || || ||

QY 238 gcacaaagactagcgaagaccatcggttgtagccgcgcgcgcgcgcgcgcgcgc 297
||||| || || || || || || || || || || || || || || || || || ||
Db 546 GGTAT-----GAAAGAGTTATCTCTCTGGCTTTGATGCATCAAGGGCTTCCAT 596
||||| || || || || || || || || || || || || || || || || || ||

QY 298 acctagctcgaatggactccgaattacgcgcgtcgactttgaacggt 348
||||| || || || || || || || || || || || || || || || || || ||
Db 597 ACCTATGCTTTCGATGGCAGCCAGGTATATTAAATGGTATGATAGCGGT 647
||||| || || || || || || || || || || || || || || || || || ||

RESULT 15
SYHA107M
LOCUS      SYHA107M      1234 bp      DNA      SYN      04-APR-2000
DEFINITION Synthetic B.macerans/B.amyloliquefaciens hybrid gene for
            endo-1,3-1,4-beta-glucanase.
ACCESSION 225874.1 GI:398187
VERSION   225874
KEYWORDS  beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE    synthetic construct.
ORGANISM  synthetic construct.
            artificial sequence.
REFERENCE 1 (bases 1 to 864)
AUTHORS   Hofemeister,J., Kurtz,A., Borriass,R. and Knowles,J.
TITLE     The beta-glucanase gene from Bacillus amyloliquefaciens shows
            extensive homology with that of Bacillus subtilis
JOURNAL   Gene 49 (2), 177-187 (1986)
MEDLINE   87192007
REFERENCE 2 (bases 1 to 1234)
AUTHORS   Borriass,R., Olsen,O., Thomsen,K.K. and von Wettstein,D.
TITLE     Hybrid bacillus endo-(1-3,1-4)-beta-glucanases: construction of
            recombinant genes and molecular properties of the gene products
JOURNAL   Carlsberg Res. Commun. 54 (2), 41-54 (1989)
MEDLINE   89374544
REFERENCE 3 (bases 463 to 852)
AUTHORS   Borriass,R., Buettner,K. and Maentsaelae,P.
TITLE     Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
            homologies to other beta-glucanases
JOURNAL   Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
MEDLINE   91109712

```

```

REFERENCE      4 (bases 1 to 1234)
AUTHORS        Politz,O., Simon,O., Olsen,O. and Borriass,R.
TITLE          Determinants for the enhanced thermostability of hybrid
            (1-3,1-4)-beta-glucanases
JOURNAL        Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE        94009045
REFERENCE      5 (bases 1 to 1234)
AUTHORS        Politz,O.
TITLE          Direct Submission
JOURNAL        Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin,
            Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
FEATURES
Source
1. .1234
/organism="synthetic construct"
/db_xref="taxon:32630"
/focus
1. .1234
/organism="Bacillus amyloliquefaciens"
/db_xref="taxon:1390"
1. .864
/organism="Bacillus amyloliquefaciens"
/db_xref="taxon:1390"
1. .1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
469. .543
/function="endo-1,3-1,4-beta-glucanase"
/notes="Hybrid between bgla (Pos. 1 to 864) and bgalm
(Pos.in hybrid 865 to 1234, Pos. original 463 to 852)"
/codon_start=1
/transl_table=11
/label="HA107M"
/evidence="experimental"
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CAA81093.1"
/db_xref="GI:398188"
/translacion="MKRVNNSMTSLGEMRLAITSYKFKDCGENRSVDTGYGLYEV
KADGYSNGNMNCTWRANNSVTSIGEMRLAITSYKFKDCGENRSVDTGYGLYEV
RMKPAKNTGIVSSFTYTGPTGTGDEIDIEFLGKDTTKVQFNYYNGVGHKVIS
LGFDAKSGFHYAFDMQPGYIKWYVDGVLKHTATANIPSTPGKIMNMLNNGTGVDDWL
GSGNANPLYAEYDMVKYTSN"
544. .1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
BASE COUNT    369 a 222 c 285 g 358 t
ORIGIN

Query Match      5.7%; Score 42.2; DB 56; Length 1234;
Best Local Similarity 54.7%; Pred. No. 0.75;
Matches 111; Conservative 0; Mismatches 83; Indels 9; Gaps 1;

QY 146 cagatggaaggccctgggtagaagtggatattgaagttctcgcgaagaatccgggcagtt 205
||||| || || || || || || || || || || || || || || || || || ||
Db 836 CGGAGGGGACTCCTGGGATGAGATTGATATCGAATTTCTAGGAAAGACACACAAAG 895
||||| || || || || || || || || || || || || || || || || || ||

QY 206 tcagtcaccaatcattaccggaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 265
||||| || || || || || || || || || || || || || || || || || ||
Db 896 TCAGTTTAACTATTATACCAATGGGTTGGCGGTTCAT-----GAAAGGTTATCT 946
||||| || || || || || || || || || || || || || || || || || ||

QY 266 ctgttagcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 325
||||| || || || || || || || || || || || || || || || || || ||
Db 947 CTCTTGGCTTTGATGATCATCAAGGGCTTCCATACCTATGCTTTTCGATTGCGACCGGTT 1006
||||| || || || || || || || || || || || || || || || || || ||

QY 326 acgtccgcgtgactgttgacggt 348
||||| || || || || || || || || || || || || || || || || || ||
Db 1007 ATATTAATGGTATGATAGCGGT 1029
||||| || || || || || || || || || || || || || || || || || ||

```

Search completed: August 23, 2001, 13:37:21
Job time: 8284 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 11:12:42 ; Search time 157.43 Seconds
(without alignments)
2967.406 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgaagattt.....gagatgacgaacctgtctcgcg 744

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.4	6.4	927	20	AA02912 B. alkalophilus be
2	44.6	6.0	783	11	AA005167 Heat-stable endo-b
3	44.6	6.0	850	11	AA003519 Thermostable beta-
4	42.2	5.7	1240	11	AA005832 Beta-glucanase-HI
5	35.2	4.7	445	21	AA071502 Neocarzinostatin a
6	34	4.6	464	21	AA071503 Neocarzinostatin a
7	34	4.6	971	19	AAV29067 Lichenase coding s
8	34	4.6	2104	22	AA087120 Nucleotide sequenc
9	33.6	4.5	1459	17	AA029043 Oerskovia beta-1,3
10	33.4	4.5	556	21	AA096632 Fusarium venenatum
11	32.8	4.4	1604	19	AAV20465 Human c-ets oncoge

12	32.8	4.4	1680	21	AA07212	P. oleoverans PHA
13	32.8	4.4	3970	14	AAQ38219	NANBH virus strain
14	32.8	4.4	4500	12	AAQ10503	6kB fragment contg
15	32.8	4.4	9589	14	AAQ38218	NANBH virus strain
16	32.8	4.4	9711	22	AA086644	Nucleotide sequenc
17	32.8	4.4	9711	22	AA086937	Nucleotide sequenc
18	32.8	4.4	9711	22	AA023486	Infectious Hepatit
19	32.8	4.4	15872	18	AA068715	Streptomyces venez
20	32.8	4.4	15872	21	AA087283	S. venezuelae vep
21	32.6	4.4	1590	21	AA080394	Human 4F2 heavy ch
22	32.6	4.4	1863	21	AA080374	Human 4F2 heavy ch
23	32.6	4.4	1897	22	AA077279	Human CDNA encodin
24	32.6	4.4	10732	21	AA010594	Gene encoding a su
25	32.2	4.3	622	21	AA091121	Fusarium venenatum
26	32.2	4.3	3401	11	AA006904	Sequence encoding
27	32	4.3	1247	21	AA013476	Aspergillus oryzae
28	31.8	4.3	1063	10	AA095000	Bacillus subtilis
29	31.8	4.3	72750	21	AA081468	N. meningitidis pa
30	31.8	4.3	34980	21	AA021544	Neisseria meningit
31	31.6	4.2	453	21	AA067724	Eucalyptus grandis
32	31.6	4.2	1473	20	AA080867	Degenerate DNA seq
33	31.6	4.2	6564	22	AA031637	Mycobacterium tube
34	31.4	4.2	691	21	AA011284	Aspergillus niger
35	31.4	4.2	987	18	AA089157	Oerskovia xanthine
36	31.4	4.2	1177	18	AA089155	Oerskovia xanthine
37	31.4	4.2	1516	18	AA089156	Oerskovia xanthine
38	31.2	4.2	601	21	AA011435	Aspergillus niger
39	31.2	4.2	667	15	AA058486	HCV peptide C14-8.
40	31.2	4.2	667	15	AA058823	NANBH virus gene f
41	31.2	4.2	3220	14	AA052535	Encodes Subtilisin
42	30.8	4.1	490	19	AA038809	Homo sapiens CESP
43	30.6	4.1	1347	22	AA060981	P. putida KT2440-a
44	30.6	4.1	5021	21	AA034703	Human tolloid-like
45	30.4	4.1	44377	18	AA078508	Platenolide syntha

ALIGNMENTS

RESULT 1
AA02912
ID AA02912 standard; DNA; 927 BP.
XX AC AA02912;
XX DT 19-MAY-1999 (first entry)
XX DE B. alkalophilus beta-glucanase coding region DNA.
XX DE Beta-glucanase; enzyme; glucan; removal; lichenan; membrane;
XX KW equipment; food industry; brewing; ds.
XX OS Bacillus alkalophilus.
XX PN DEL9732751-AL.
XX PD 04-FEB-1999.
XX PF 30-JUL-1997; 97DE-1032751.
XX PR 30-JUL-1997; 97DE-1032751.
XX (HENK) HENKEL KGAA.
XX PI Hillen W, Maurer K;
XX XX WPI; 1999-122161/11.
XX DR P-PSDB; AAW93001.
XX PT New Bacillus alkalophilus beta-glucanase protein and gene - useful
XX PT for removing glucan and/or lichenan from membranes in the brewing
XX PT industry

DR P-PSDB; AAR03775.

XX Thermotable beta-glucanase production -
PT using Bacillus subtilis transformed with gene from Bacillus
PT macerans.

XX Fig 1; ; 9pp; German.

XX The gene encoding Bacillus macerans thermostable
CC beta-glucanase is expressed in Bacillus subtilis. The
CC enzyme is useful for lowering the viscosity of brewing
CC mashes and in the production of feedstuff.

XX Sequence 850 BP; 250 A; 160 C; 196 G; 244 T; 0 other;
SQ

Query Match 6.0%; Score 44.6; DB 11; Length 850;

Best Local Similarity 52.9%; Pred. No. 0.00043;
Matches 154; Conservative 0; Mismatches 119; Indels 18; Gaps 2;

QY 58 tacggttaagttaagcccgatgaagatggcagccgcatcggaacagtcagttccatg 117

DB 373 tacggcgtgacgaggtcagtagaagccagcccaataacaggaattgtctcattctt 432

QY 118 ttctctaccagaatggttcgaaatcccgatggcagcccggtgtagaagtgatatt 177

DB 433 ttacgta-----tacagacctgctcatggcacaataatgggatgaatagatc 483

QY 178 gaagtctcgcaagaatccggcgagttccagtcacacattaccgtaagccggc 237

DB 484 gaatttcagaaagacagcaaaagtcagtttaactattatataccaatgggttgcg 543

QY 238 gcacaaagactagcgaagcaccatgctgttagcccgcccgatccaggtttccac 297

DB 544 ggtcat-----gaaaggttattctctctgtggttgatcatcaagggttccat 594

QY 298 acctacggttcgaatggaactcgaattacgtccgctggactgttgacggt 348

DB 595 acctatgcttcgattggcagccggtatataataatggatgtgacggt 645

RESULT 4

AAQ05832
ID AAQ05832 standard; DNA; 1240 BP.

XX AC AAQ05832;

XX 09-JAN-1991 (first entry)

XX Beta-glucanase-H1 gene.

XX Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry; ss.

XX Bacillus amyloliquefaciens, Bacillus macerans.

XX Key Location/Qualifiers

FT CDS 469..1188

FT /*tag= a

FT /label=Hybrid pre-glucanase

FT misc_RNA 469..864

FT /*tag= b

FT /label=N-terminal of B.amyloliquefaciens beta-

FT glucanase

FT misc_RNA 865..1185

FT /*tag= c

FT /label=C-terminal half of B.macerans beta-

FT glucanase

XX WO9009436-A.

XX PN

XX 23-AUG-1990.

XX PD

XX 16-FEB-1990; 90WO-DK00044.

XX PF

XX

XX

PR

PR

XX

PA

PA

XX

PI

XX

DR

DR

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity

Matches 111; Conservative

QY 146 ccgatgaagccctggtagaagtgaattgaattcgcgaagaatccggcgagtt 205

DB 836 cggaggggactcctgggtagagattgatactcgaattcttaggaagacacacaaag 895

QY 206 tccagtcacaacatcattaccgggtaaggccgcccacacaaagactagcgaagacaccatg 265

DB 896 tccagtttaactattataccaatgggttggtggtcat-----gaaaaggttatct 946

QY 266 ctgttagcccgccgcatcaggctttccacacctacgctcgaatggactccgaatt 325

DB 947 ctctggctttgatgcatacaagggttccataacctatgcttcttgatggcagccaggt 1006

QY 326 acgtccgctgactgttgacggt 348

DB 1007 atataaatggtatgtagacggt 1029

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC of grain (e.g. barley or wheat) in the brewing and fermentation
CC industries to increase carbon substrate availability and, to maximise
CC production of desired products.

SQ Sequence 971 BP; 373 A; 117 C; 155 G; 326 T; 0 other;
SQ Sequence 971 BP; 373 A; 117 C; 155 G; 326 T; 0 other;

Query Match	Score 34:	DB 19:	Length 971:
4.68:	Score 34:	DB 19:	Length 971:

Best Local Similarity 51.4%;
Matches 110; Conservative 0; Mismatches 95; Indels 9; Gaps 1;

Qy 148 gatggaaggccctgggtagaagtggatattgaagttctcggccaagaatccgggcagtttc 207

Db 504 gatggaactaagtgggatgaaattgatatagaattcccttqqtattatgatacaaccaaattt 563

Qy 208 cagtccaacatcattaccggtaaqqccqccacaaaqaactaqqcaaaacaccatqct 267

Db 564 caatttaactactacactaat-----ggacaaaggtcatcatgaacataatcattat 614

Qy 268 gttagccccgcgcgcgatcagggtttccacacctacggtctcgaatggactccaattac 327

Db 615 cttggatttgatgcctctcaaggattccatacctatggtttcttctggcgagaattct 674

QY 328 gtccgctggactatttgacagctcaggaagtccgca 361

Db 675 attacatggtatgtagatggtacagcgtttaca 708

RESULT 8

AAC87120

ID AAC87120 standard; cDNA: 2104 bp.

AAC87120;

DT 20-APR-2001 (first entry)

DE Nucleotide sequence of a plant auxin transporter polypeptide.

KW Auxin: auxin transporter: AUX1: root gravitropism: transgenic plant:

KW plant phenotype; herbicide; ss.

OS *Oryza sativa*.

FH	Key	Location/Qualifiers
100	100	100
101	101	101
102	102	102
103	103	103
104	104	104
105	105	105
106	106	106
107	107	107
108	108	108
109	109	109
110	110	110
111	111	111
112	112	112
113	113	113
114	114	114
115	115	115
116	116	116
117	117	117
118	118	118
119	119	119
120	120	120
121	121	121
122	122	122
123	123	123
124	124	124
125	125	125
126	126	126
127	127	127
128	128	128
129	129	129
130	130	130
131	131	131
132	132	132
133	133	133
134	134	134
135	135	135
136	136	136
137	137	137
138	138	138
139	139	139
140	140	140
141	141	141
142	142	142
143	143	143
144	144	144
145	145	145
146	146	146
147	147	147
148	148	148
149	149	149
150	150	150
151	151	151
152	152	152
153	153	153
154	154	154
155	155	155
156	156	156
157	157	157
158	158	158
159	159	159
160	160	160
161	161	161
162	162	162
163	163	163
164	164	164
165	165	165
166	166	166
167	167	167
168	168	168
169	169	169
170	170	170
171	171	171
172	172	172
173	173	173
174	174	174
175	175	175
176	176	176
177	177	177
178	178	178
179	179	179
180	180	180
181	181	181
182	182	182
183	183	183
184	184	184
185	185	185
186	186	186
187	187	187
188	188	188
189	189	189
190	190	190
191	191	191
192	192	192
193	193	193
194	194	194
195	195	195
196	196	196
197	197	197
198	198	198
199	199	199

FT	CDS	
250..1728		

```
FT /*tag= a
```

PN WO200078965-A2.

PD 28-DEC-2000.

21-JUN-2000: 2000WO-US17050.

22-JUN-1999: 99US-0140212.

PA (DUPO) DU PONT DE NEMOURS & CO E T.

PI Cahoon RE, Weng Z:

WPI: 2001-071396/08.

XX
F-300B, HARB312339.

PT phytohormone auxin. useful for screening for protein inhibitors for

PT potential as herbicides -

PS Claim 2; Page 51; 69pp; English.

CC The present sequence encodes an auxin transporter. The polypeptide
CC mediates he effects if the plant hormone auxin. The polypeptides of
CC the invention are homologous to the Arabidopsis auxin transporter AUX1,
CC which effects root gravitropism. The polynucleotides are used for

PT useful for modifying or degrading beta-glucan contg. material and in
 PT the prepn. of e.g. food colourants, flavourings and yeast extracts
 XX
 PS
 XX Claim 1; Page 41-43; 60pp; English.

CC A genomic DNA sequence (AAT29043) codes for a novel beta-1,3-
 CC glucanase (AAR97362) useful for degrading or modifying beta-glucan-
 CC contg. material. It was obt'd. from a genomic DNA library of
 CC Oerskovia xanthineolytica LG109 by PCR amplification, screening
 CC of chromosomal DNA using a PCR product as probe, prepn. of a
 CC gene bank in pUC18, and recovery of plasmid pPR8A which contained
 CC the gene as a 1.5 kb BamHI-kpnI fragment. The beta-1,3-glucanase
 CC gene can be used for large-scale prodn. of recombinant enzyme, esp.
 CC using Bacillus subtilis as host. Protease-free enzyme is obt'd.
 CC which can be used to lyse fungal cell walls, for protoplast prepn.,
 CC or for prodn. of pigments, colorants, flavour, yeast extract and
 CC pharmaceuticals.

XX Sequence 1459 BP; 254 A; 503 C; 475 G; 227 T; 0 other;

Query Match 4.5%; Score 33.6; DB 17; Length 1459;

Best Local Similarity 66.7%; Pred. No. 1.6;

Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 282 cgatcagggttccacacctaagtcggtcgaatggactccgaattacgtccgctgactgt 341

DB 1021 cggcagcagctccacacattcggcatcgtggtgacgcgggtgagatcacgtgctcgt 1080

QY 342 tgacggtcagga 353

DB 1081 tgacgggcagga 1092

RESULT 10

AAFO9632

ID AAF09632 standard; cDNA; 556 BP.

AC AAF09632;

DT 13-MAR-2001 (first entry)

DE Fusarium venenatum EST SEQ ID NO:2155.

XX Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Fusarium venenatum.

OS WO200056762-A2.

PN 28-SEP-2000.

PD 22-MAR-2000; 2000WO-US07781.

PF 22-MAR-1999; 99US-0273623.

PR (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX WPI; 2000-594572/56.

DR Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a

PT substrate of expressed sequence tags

XX Claim 86; Page 1182; 3161pp; English.

PS
 XX

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

XX Sequence 556 BP; 147 A; 150 C; 126 G; 132 T; 1 other;

Query Match 4.5%; Score 33.4; DB 21; Length 556;

Best Local Similarity 51.0%; Pred. No. 1.2;

Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 588 ccgctgggggcaagggtgactggacatgttgacggttaacggtgcacctcacgcacaagaa 647

DB 369 cccctggccacggtatataaaccttcacctggatgtaccacatcagcagcaacac 428

QY 648 catctactccagagatggcatgttgatctcgcctccctcacccgaaaggtcaggaaagctt 707

DB 429 ccacggctgtagtgcatacaagacctcttcgacatcagctgcagatgctataagaggttg 488

QY 708 caacggccagggtccgagagatgacgaacctctc 742

DB 489 gaagggtgctgattggcactgcgacgactcgtcgc 523

RESULT 11

AAV20465

ID AAV20465 standard; DNA; 1604 BP.

XX AAV20465;

XX 17-JUN-1998 (first entry)

DT Human c-ets oncogene.

DE Human; Oncogene; proto-oncogene; neoplastic disease; anticancer;

KW cancer; antisense oligonucleotide; c-ets; ds.

XX Homo sapiens.

OS US5734039-A.

PN 31-MAR-1998.

PD 15-SEP-1994; 94US-0306691.

PF 15-SEP-1994; 94US-0306691.

PR (UJJE-) UNIV JEFFERSON THOMAS.

PA Calabretta B, Skorski T;

XX WPI; 1998-229882/20.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 11:12:42 ; Search time 1874.7 Seconds
(without alignments)
3751.493 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgctcg 744

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_estl1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est33:*
26:	gb_est34:*
27:	gb_est35:*
28:	gb_est36:*
29:	gb_est37:*
30:	gb_est38:*
31:	gb_est39:*
32:	gb_est40:*
33:	em_estba:*
34:	em_estfun:*
35:	em_esthum1:*
36:	em_esthum2:*
37:	em_esthum3:*
38:	em_esthum4:*
39:	em_esthum5:*
40:	em_esthum6:*
41:	em_esthum7:*
42:	em_esthum8:*
43:	em_esthum9:*

44:	em_esthum10:*
45:	em_esthum11:*
46:	em_esthum12:*
47:	em_esthum13:*
48:	em_esthum14:*
49:	em_esthum15:*
50:	em_esthum16:*
51:	em_esthum17:*
52:	em_esthum18:*
53:	em_esthum19:*
54:	em_esthum20:*
55:	em_esthum21:*
56:	em_esthum22:*
57:	em_esthum23:*
58:	em_esthum24:*
59:	em_esthum25:*
60:	em_esthum26:*
61:	em_esthum27:*
62:	em_esthum28:*
63:	em_estin1:*
64:	em_estin2:*
65:	em_estin3:*
66:	em_estin4:*
67:	em_estin5:*
68:	em_estomi:*
69:	em_estom2:*
70:	em_estov1:*
71:	em_estov2:*
72:	em_estpl1:*
73:	em_estpl2:*
74:	em_estpl3:*
75:	em_estpl4:*
76:	em_estpl5:*
77:	em_estpl6:*
78:	em_estpl7:*
79:	em_estpl8:*
80:	em_estpl9:*
81:	em_estpl10:*
82:	em_estro1:*
83:	em_estro2:*
84:	em_estro3:*
85:	em_estro4:*
86:	em_estro5:*
87:	em_estro6:*
88:	em_estro7:*
89:	em_estro8:*
90:	em_estro9:*
91:	em_estro10:*
92:	em_estro11:*
93:	em_estro12:*
94:	em_estro13:*
95:	em_estro14:*
96:	em_estro15:*
97:	em_estro16:*
98:	em_estro17:*
99:	em_estro18:*
100:	em_estro19:*
101:	em_estro20:*
102:	gb_est25:*
103:	gb_est26:*
104:	gb_est27:*
105:	gb_est28:*
106:	gb_est29:*
107:	gb_est30:*
108:	gb_est31:*
109:	gb_est32:*
110:	gb_est41:*
111:	gb_est42:*
112:	gb_est43:*
113:	gb_est44:*
114:	gb_est45:*
115:	gb_est46:*
116:	gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	40.8	5.5	418	136	BE495688	WHE1279_B
2	37.8	5.1	796	15	AI068991	mgae0005a
3	37.8	5.1	849	15	AI069170	mgae0005d
4	37.2	5.0	561	173	BG103002	RHI22_35
5	37	5.0	552	19	AI397807	NCM9C4T3
6	36.8	4.9	747	170	BF861648	963025E08
7	36.4	4.9	604	166	BE359676	DG1_56_H0
8	36.4	4.9	608	154	BG488141	RHI22_60
9	36.2	4.9	635	167	BE405426	WHE1216_C
10	36	4.8	599	155	BG605030	D
11	35.8	4.8	387	145	BF202533	WHE1776_G
12	35.8	4.8	584	13	AA898069	NCM3G7T7
13	35.8	4.8	587	13	AA897960	NCM1H9T7
14	35.4	4.8	640	141	BE888160	601511671
15	35.4	4.8	669	164	BE208885	GF-FV-P3C
16	35.2	4.7	776	151	BF627470	HVSMED000
17	35	4.7	340	116	AW486705	76371_MAR
18	35	4.7	360	167	BE428429	MTD006_H1
19	35	4.7	423	167	BE398687	WHE0024_C
20	35	4.7	482	166	BE347923	sp08h02_Y
21	35	4.7	495	162	BE059962	sn39c06_Y
22	35	4.7	762	170	BF865955	963063D08
23	34.8	4.7	781	141	BE884266	601505733
24	34.6	4.7	826	146	BF263101	HV_CEA000
25	34.6	4.7	876	150	BF526437	602070921
26	34.4	4.6	443	104	AI940919	sb1a11_Y
27	34.4	4.6	474	32	AV668531	AV668531
28	34.4	4.6	529	136	BE494282	WHE1253_H
29	34.4	4.6	536	149	BF473353	WHE0923_E
30	34.4	4.6	552	167	BE419465	WWS012_D8
31	34.2	4.6	376	144	BF072772	NCSW4G11T
32	34.2	4.6	455	175	BG278686	ae05np_r
33	34.2	4.6	501	175	BG279070	af05np.f
34	34.2	4.6	521	175	BG279385	b2g07np_r
35	34.2	4.6	532	175	BG279532	b3g11np_r
36	34.2	4.6	546	30	AV386476	AV386476
37	34.2	4.6	548	13	AA898699	NCM6G5T7
38	34.2	4.6	550	13	AA898546	NCP5G127
39	34.2	4.6	552	175	BG279542	b3h07np.f
40	34.2	4.6	558	13	AA898246	NCM4A9T7
41	34.2	4.6	560	175	BG280207	glf11np.f
42	34.2	4.6	564	175	BG278321	a3e03np.f
43	34.2	4.6	564	175	BG279488	b3e07np.f
44	34.2	4.6	569	13	AA898089	NCM3D11T7
45	34.2	4.6	584	13	AA897951	NCM2B12T7

ALIGNMENTS

Result	1
LOCUS	BE495688 418 bp mRNA EST
DEFINITION	WHE1279_B01_C012S Secale cereale anther cDNA library Secale cereale
ACCESSION	BE495688
VERSION	BE495688.1 GI:9662281
KEYWORDS	EST.
SOURCE	rye.
ORGANISM	Secale cereale
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Secale.
AUTHORS	1 (bases 1 to 418)
TITLE	Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Ross, K., Seaton, C.L. and Tong, J.C.

TITLE	The structure and function of the expressed portion of the wheat genomes - Anther cDNA library from rye
JOURNAL	Unpublished (2000)
COMMENT	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Email: oanderson@pw.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Stratagene SK primer.
FEATURES	Location/Qualifiers
source	1. 418
	/organism="Secale cereale"
	/cultivar="Blanco"
	/db_xref="taxon:4550"
	/clone="WHE1279_B01_C01"
	/tissue_type="Anther"
	/dev_stage="Adult plant before anthesis"
	/lab_host="E. coli SOLR"
	/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site1: EcoRI; Site2: XhoI; Plants were grown in the greenhouse. Anthers were harvested and pooled from early meiosis to late meiosis. The tissue, total RNA, and poly(A) RNA were prepared (Butler, Ross and Gustafson) at University of Missouri, Columbia. A cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT	77 a 180 c 103 g 58 t
ORIGIN	
Query Match	5.5%; Score 40.8; DB 136; Length 418;
Best Local Similarity	56.8%; Pred. No. 0.11;
Matches	75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
Qy	583 ggctccgcctgggcaagggtgactgacatttgacggtaacccgtgctgcacctcacgac 642
Db	90 GCCTCCTCCTTGTCTGACGACAAAGGGCGGAGGACGCGATCCGGGTCTCTCCGCGAG 149
Qy	643 aagaaatctactcagagatggcatgttgatcctccctcacgcgcaaaaggtcagaa 702
Db	150 AAAAGGATCGACCTACCGGCGCCCATCGTCAAGGTCTTCTCACCCCTCAAGGGGAGAAC 209
Qy	703 agcttcaacggc 714
Db	210 GCCGCCACCGGC 221
RESULT	2
LOCUS	AI068991 796 bp mRNA EST 09-DEC-1999
DEFINITION	mgae0005B09f Magnaporthe grisea Appressorium Stage cDNA Library
ACCESSION	AI068991
VERSION	AI068991.1
KEYWORDS	GI:3391966
SOURCE	EST.
ORGANISM	Magnaporthe grisea.
REFERENCE	Magnaporthe grisea Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
AUTHORS	1 (bases 1 to 796)
TITLE	Choi, W., Fang, E., Sasinowski, M., Wing, R. and Dean, R.A.
JOURNAL	Expressed sequence characterization during appressorium formation in rice blast fungus, Magnaporthe grisea
COMMENT	Unpublished (1998) Contact: Dean, R.A.

	DATE	COUNT	ORIGIN
✓	127	5	
✓	137	5	
✓	172	9	
✓	110		
✓	110		

phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

0987
QY

QY 286 cagggcttccacacccatcagggctcgcgaatcgacccgaattacgctccgctcggactgcttgac 34

Best Local Similarity 54.5%; Pred. No. 3.8;
Matches 72; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 583 ggcctccgctgagggaaggtgacttgacatttgacggttaacggtgtgacactcaccgac 642
Db 83 GGCTCACCATTGACAAAGGAGGCGACCACTAGCAGGGAAGCGCGCGCGCCGCAAC 142
QY 643 agaaatctactccagagatggcatgttgatctcctcaccgcgaaggtcagga 702
Db 143 GCAACAGCCTCTCAGCGCGCGGGTTCATCCCTCGCAGCACCAACACCCCGGAG 202
QY 703 agcttcacggc 714
Db 203 GGCAGCGCGGC 214

RESULT 11
BF202533
LOCUS
DEFINITION
aestivum cDNA clone WHE1776_G11_M22, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum
broad wheat.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.

REFERENCE
AUTHORS
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
Seaton, C.L. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Pre-anthesis spike cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959573
Fax: 5105959581
Email: oanderson@nwp.usda.gov

Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
Location/Qualifiers
1. 387
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1776_G11_M22"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ Close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

BASE COUNT
ORIGIN
86 a 123 c 125 g 53 t
Query Match
Best Local Similarity 61.1%; Pred. No. 3.9;
Matches 58; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

FEATURES
source

1. 387
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1776_G11_M22"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ Close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

BASE COUNT
ORIGIN
86 a 123 c 125 g 53 t
Query Match
Best Local Similarity 61.1%; Pred. No. 3.9;
Matches 58; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 628 gtcgacctcaccgacaagaacatctactccagagatggcatgttgatcctcgccctcacc 687
Db 201 GCGGACCTCACCCTTACAGAACTGGATCGAGTGTGCGCGTGGACAACTTCACC 260
QY 688 gcaagggtcagaaagcttcaacgcccagggtcc 722
Db 261 ATCAAGCGGCACGGCACCACATCGAGGCCAGGGCCC 295

RESULT 12
AA898069/c
LOCUS
DEFINITION
Myccelial Neurospora crassa cDNA clone NM3G7 3' end, mRNA
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Neurospora crassa
Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 584)
Neelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,
Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueves, E.,
Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K., Miller, R.,
Ortega, J., Pavlova, I., Perea, J., Todisco, S., Trujillo, R.,
Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S. and Natvig
D.O.
Expressed sequences from conidial, mycelial, and sexual stages of
Neurospora crassa
Fungal Genet. Biol. 21, 348-363 (1997)
97435549
Contact: Natvig, D.O./Nelson, M.A.
Department of Biology
University of New Mexico
Castetter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304
Email: ngp@biology.unm.edu
Deposited in GDB as S:1146980
accession GDB:S:1146980
Seq primer: T7.

FEATURES
Location/Qualifiers
1. 584
/organism="Neurospora crassa"
/strain="74-OR23-IV A (F6SC 2489)"
/db_xref="taxon:5141"
/clone="NM3G7"
/clone_lib="Myccelial"
/sex="Mating type A"
/tissue_type="Mycelium"
/dev_stage="Mycelium"
/lab_host="E. coli"
/note="Vector: pBluescript SK (-); Site 1: EcoR I; Site 2:
Xho I; 2% sucrose for 24 hours. cDNA directionally cloned
into pBluescript SK (-) using the Uni-ZAP XR vector system
(Stratagene, La Jolla, CA)."
132 a 181 c 122 g 149 t

BASE COUNT
ORIGIN
132 a 181 c 122 g 149 t
Query Match
Best Local Similarity 4.8%; Score 35.8; DB 13; Length 584;
Matches 115; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 447 tggccagttcgaatcaaaagcttccggttttcaggttcacaaactgggtcaagttta 506
Db 521 TGGTGCCTTCGAAGACTACGTGATCCGATCGACAAGCGTCACACTCGTCGAGGACTA 462
QY 507 taagtataccgcccggcggaagggcgagcagcttactgactgagcagcaaa 566
Db 461 CGCATCGGGGTCGAACAGCGTGGTGGCGTCGAGGACTATGTGATCGAGTCGACAAACG 402


```

Db      395  GGACGGTATCTTGATAAGAG 375
|||||
RESULT 15
BE208885
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE208885 669 bp mRNA 28-JUN-2000
GF:FPV-P3C5 Marsh grapefruit young flavedo Citrus x paradisi cdna,
mRNA sequence.
BE208885
BE208885.1 GI:8789137
EST.
Citrus x paradisi.
Citrus x paradisi.
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 669)
McCollum, T.G., Doostdar, H., Shatters, R. and Mayer, R.T.
Marsh grapefruit young flavedo EST
Unpublished (2000)
Contact: McCollum, T.G.
US Horticultural Research Laboratory
USDA, ARS
2001 S. Rock Road, Ft. Pierce, FL 34945, USA
Tel: 561-462-5836
Fax: 561-462-5986
Email: gmccollum@ushrl.ars.usda.gov
Plate: 3 row: C column: 5.
Location/Qualifiers
1. 669
/organism="Citrus x paradisi"
/cultivar="Marsh"
/db_xref="taxon:37656"
/clone_lib="Marsh grapefruit young flavedo"
/tissue_type="Flavado"
/dev_stage="2-5 gram fruit"
/note="Organ: Fruit"
163 a 138 c 187 g 181 t

BASE COUNT
ORIGIN

Query Match 4.8%; Score 35.4; DB 164; Length 669;
Best Local Similarity 51.6%; Pred. No. 6.1;
Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 207 ccagtcacaatcattaccggtaaggccggcgacacaaagactagcgaaagcaccatgc 266
||| |||||
Db 41 CGGTACCAAGCAAACTGCACCTAAGTCAACCGGAGGAAGGCTCTTAGGAAGCAACTTGC 100
||| |||||
QY 267 tgttagcccccgcgcgcgcgcgcgttctccacacctacggtctcgaaatggactccgaatta 326
||| |||||
Db 101 TACCAAGGCCGCCGTAGTCTGCGGCCCAACCAACCGGGTGGTGTAAAGACCCATCGTTA 160
||| |||||
QY 327 cgtccgcgtggaactgttgacggtcagggaagtcgcgaag 363
||| |||||
Db 161 CGGTCTGTGAACCGCTGCGCCCTTCGAGAAATTCGCAAG 197
||| |||||

```

Search completed: August 23, 2001, 12:48:29
Job time: 5747 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:14:13 ; Search time 2906.88 Seconds
(without alignments)
3958.886 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgtctcg 744

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues
Word size: 0
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_in4:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_vil2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pri:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_rol:*

95: gb_rol2:*

96: gb_in4:*

97: gb_pri10:*

98: em_ba3:*

pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	741	99.6	1426	3	FIBGLUC
2	20	2.7	1289	12	AB009683
3	20	2.7	8435	5	AF236106
4	20	2.7	70841	4	AC004287
5	20	2.7	93588	64	AC017187
6	20	2.7	268219	5	AE003806
7	19	2.6	1944	97	HSU45983
8	19	2.6	2608	97	HSU62556
					M33676 F.succinoge
					AB009683 Raphanus
					AF236106 Drosophil
					AC004287 Drosophil
					AC017187 Drosophil
					AE003806 Drosophil
					U45983 Homo sapien
					U62556 Human chemo

```
c 9 2.6 2608 97 HSU62556 U62556 Human chemo
c 10 2.6 86172 91 AY016370 Homo sapi
c 11 2.6 86172 91 AY016370 Homo sapi
c 12 2.6 98861 89 AL138961 Human DNA
c 13 2.6 147009 76 AC084179 Homo sapi
c 14 2.6 151174 64 AC016189 Homo sapi
c 15 2.6 159852 70 AC026309 Homo sapi
c 16 2.6 159852 70 AC026309 Homo sapi
c 17 2.6 166826 80 AL356535 Homo sapi
c 18 2.6 167506 69 AC025679 Homo sapi
c 19 2.6 178563 83 AP002419 Homo sapi
c 20 2.6 194014 81 AL583784 Homo sapi
c 21 2.6 213524 61 AC010735 Homo sapi
c 22 2.6 220055 74 AC069499 Homo sapi
c 23 2.6 86 9 AR045081 Homo sapi
c 24 2.6 86 10 I18613 Sequence 82
c 25 2.6 86 10 I34157 Sequence 82
c 26 2.6 702 54 G51819 SHGC-79264
c 27 2.6 1213 95 RATGSTPPS M14364 Rat glutath
c 28 2.6 1213 94 AF156988 Rattus no
c 29 2.6 1352 94 AF156989 Rattus no
c 30 2.6 1381 1 AF037600 Rattus no
c 31 2.6 1532 3 STMSALO D50081 Streptomyce
c 32 2.6 3602 14 CPSUCPSY2 Y11795 C.plantagin
c 33 2.6 5591 1 AF027499 AF027499 Azotobact
c 34 2.6 7425 2 AX024365 Sequence
c 35 2.6 7425 9 AX024258 Sequence
c 36 2.6 7805 65 AC018084 Sequence
c 37 2.6 10145 1 AE000839 AC018084 Drosophil
c 38 2.6 13416 1 AE000117 AE000839 Methanoba
c 39 2.6 34074 3 SCF81 AE000117 Escherich
c 40 2.6 34766 3 SC121 AL133171 Streptomy
c 41 2.6 37931 3 SC101 AL137166 Streptomy
c 42 2.6 41565 6 AL359988 Streptomy
c 43 2.6 47713 2 CX57A10B AL032847 Caenorhab
c 44 2.6 47713 2 AX024320 Caenorhab
c 45 2.6 47887 63 AX024213 Sequence
c 46 2.6 47887 63 AC015109 Drosophil
```

ALIGNMENTS

```
RESULT 1
FIBGLUC 1426 bp DNA BCT 26-APR-1993
LOCUS FIBGLUC 1426 bp DNA BCT 26-APR-1993
DEFINITION F. succinogenes 1,3-1,4-beta-D-glucan 4-glucanohydrolase gene,
complete cds.
ACCESSION M33676 M33311
VERSION M33676.1 GI:148575
KEYWORDS 1,4-beta-D-glucan 4-glucanohydrolase; mixed-linkage beta-glucanase.
SOURCE F. succinogenes (strain S85) DNA, clone PJ15.
ORGANISM Fibrobacter succinogenes
Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;
Fibrobacter.
REFERENCE 1 (bases 1 to 1426)
AUTHORS Teather, R.M. and Erfile, J.D.
TITLE DNA sequence of a Fibrobacter succinogenes mixed-linkage
beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene
JOURNAL J. Bacteriol. 172, 3837-3841 (1990)
MEDLINE 90299807
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by R.M. Teather, 11-APR-1990.
FEATURES
source
location/Qualifiers
1..1426
/organism="Fibrobacter succinogenes"
/db_xref="taxon:833"
35_signal 62..66
-10_signal 85..90
RBS 132..137
/note="ribosome binding site"
145..1194
CDS /note="1,3-1,4-beta-D-glucan 4-glucanohydrolase precursor
(EC 3.2.1.73)"
```

```
/codon_start=1
/transl_table=11
/protein_id="AAA24896.1"
/db_xref="GI:148576"
/translation="MNKKTAVKSAALAAATTTNWSAKDFSGABLYLTLEVOYQK
FEARMKMAASGVSSMFLYONGSEIADGRPWVEVIEVLGNKPGSNITGKAG
AKTSEKHAVSPAADAFHTYGLWTFYVWTVGQVGRKTEGQVSNLTGTGLR
FNLSSESAAWVGQDFESKLPLOF INNVKVKYTPGQEGSGDFTLDWTFDGLR
SRWCKGWTFDGNVDLTDKNIXSRDGLILALFKQGESFNGQVPRDEPAQSSSS
APASSSVPPASSSVPPASSSAFAFPSSSSATNAIHGMRITTPAVAKEHRLNVAKGAK
VNPNGHKRYRVNFEH"
sig_peptide 145..225
/note="1,3-1,4-beta-D-glucan 4-glucanohydrolase signal
peptide"
mat_peptide 226..1191
/note="1,3-1,4-beta-D-glucan 4-glucanohydrolase"
BASE COUNT 371 a 346 c 335 g 374 t
ORIGIN
Query Match 99.6%; Score 741; DB 3; Length 1426;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;
Matches 741; Conservative 0;
Qy 4 gttagcgaagatttttagcggtggtccgaactctcacagttagaagaagttcagtcaggt 63
|||||
Db 217 GTTACGCGAAAGGATTTAGCGGTCCGCACTCTACACGTTAGAGAGAGTTTCCAGTACGGT 276
Qy 64 agtttgagccggtatgaagatggcagcgcatcgggaacagtcagttccatgtctctc 123
|||||
Db 277 AAGTTTGAAGCCGCTATGAAGATGGCAGCGCATCGGGAACAGTCAAGTTCATGTTCTCTC 336
Qy 124 taccagaatgttcgaatcgcgatggaagcctgggtagaagttgagattgaagtt 183
|||||
Db 337 TACCAGATGGTTCCGAATCGCCGATGGAGGCCCTGGGTAGAAGTGGATATTGAAGTT 396
Qy 184 ctggcgaagatccggcgagttctccagtcacacatcatcaccgtaagcgcgcgacacaa 243
|||||
Db 397 CTCGCGAAGATCCGGGCGAGTTTCCAGTCCACATCATATTACCGGTAAAGCGCGCGCACAA 456
Qy 244 agactagcgaagacacacatgctgttagcccgccgcatcaggtcttcacacacctac 303
|||||
Db 457 AAGACTAGCGAAAGACCATGCTGTAGCCCGCCGCGATCAGGCTTTCACACACCTAC 516
Qy 304 ggtctcgaatggatccggaattacgtccgctggaactgttgacggtcaggaagatccgcaag 363
|||||
Db 517 GGTCTCGAATGGACTCCGAAATTTACGCTCGGCTGGAGTGTGACGGTTCAGGAAGTCCGCAAG 576
Qy 364 acggaagtgccaggtttccaaactgcaggtacacagggactccgttttaacaccttgg 423
|||||
Db 577 ACGGAAGTGGCCAGGTTTCCAACTTGACAGGTACACAGGGACTCCGTTTAACTTTTGG 636
Qy 424 tcgtctgagtgccggttgggtggccagttcgatgaatcaaaagcttcgcgttttccag 483
|||||
Db 637 TCGTCTGAGAGTGGCGCTTGGTTCGCCAGTTTCGATGAATCAAAAGCTTTCGCTTTTCCAG 696
Qy 484 ttcataactgggttcgaagtttataagttacacgcggcgccagggcgagggcgagcgac 543
|||||
Db 697 TTCATCAACTGGGTCAAGGTTTATAAGTATACGCCGGGCCAGGGCGAAGGGCGAGCGAC 756
Qy 544 tttaacgcttgactgacgcgaatatttgaacagtttgatgctcccgctggggcgaaggt 603
|||||
Db 757 TTTACGCTTGACTGGACCGACAAATTTTACACAGTTTGTGATGGCTCCCGTGGGCGAAGGTT 816
Qy 604 gactggacatttgacggttaaccggtgctgacacctcaccgcaagaatactactccagagat 663
|||||
Db 817 GACTGGACATTTGACGGTAAACCGTGTGCACCTCACCCGACAAAGAACATCTACTCCAGAGAT 876
Qy 664 ggcattgtgactcctgcctcaccgcgaagaaagtcaggaagaaagcttcaacggcaggttcgcg 723
|||||
Db 877 GGCATGTTGATCCTCGCCCTCACCGCAAGGTCAGAAAGCTTCAACGGCGCAGGTTCCG 936
Qy 724 agagatgacgaacctgctccg 744
```

```
|||||
Db 937 AGATGACGAACCTGCTCCG 957

RESULT 2
LOCUS AB009683 1289 bp DNA PLN 25-JUL-1998
DEFINITION Raphanus sativus SLG(S8) gene for S glycoprotein, partial cds.
ACCESSION AB009683
VERSION AB009683.1 GI:3327851
KEYWORDS SLG(S8); S glycoprotein.
SOURCE Raphanus sativus DNA.
ORGANISM Raphanus sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Raphanus.
1 (bases 1 to 1289)
Kusaba, M. and Nishio, T.
Direct Submission
Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Makoto
Kusaba, National Institute of Agrobiological Resources, Institute
of Radiation Breeding; 2425 Kamimurata, Ohmiya, Ibaraki 319-22,
Japan (E-mail:Kusaba@irb.affrc.go.jp, Tel.81-2955-2-1138,
Fax:81-2955-3-1075)
2 (sites)
Sakamoto, K., Kusaba, M. and Nishio, T.
Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus
related gene (SLR1) in Raphanus sativus L. and self-incompatible
ornamental plants in the Brassicaceae
Mol. Gen. Genet. 258 (4), 397-403 (1998)
JOURNAL 98311079
MEDLINE
FEATURES
source Location/Qualifiers
1..1289
/organism="Raphanus sativus"
/db_xref="taxon:3726"
<1..>1289
/gene="SLG(S8)"
/note="PCR product"
/product="S glycoprotein"
1..1289
/gene="SLG(S8)"
<1..>1289
/gene="SLG(S8)"
/codon_start=2
/product="S glycoprotein"
/protein_id="BAA31730.1"
/db_xref="GI:3327852"
/translation="DSSYTLFLLVFFVLLFGPAHSINTLSSTESLTSSNRTLLSP
GNVFEGLFPKSSRWYLGWTKLPDRYVWVANKDPLSNISGTLKISNNLVLL
DHSKSWNTLNRGRSPVVAELLANGNFVTRYNNNDASEFLWQSDFFDTLLP
EMKGLDLAKGLNRLTSWKNDDPSSEISYQLDIQRTPEFLKDGFRGYNRPW
NGVRNGIPEDQNLVWYNFTDSEVAVTFLTDKNISYSLIISNDEYLA RLTLTP
ASDWNLFTSPSEPCDVYMTGCPVAYCDVNTSPVCNIOGKPLNVQVQDLRDGSG
GCIRTELSGSDGFRFMKNKLPDTHMTATVDRSICGVKCEKRLSDCNCATAPANADV
RNGGTGCVIWTGKLDIRNYFFDQGLVRLAAADLV"
BASE COUNT 349 a 267 c 313 g 360 t
ORIGIN
Query Match 2.7%; Score 20; DB 12; Length 1289;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 accgacaagaactctactc 656
|||||
Db 758 ACCGACAGACATCTACTC 777

RESULT 3
LOCUS AF236106 8435 bp mRNA INV 01-MAR-2000
DEFINITION Drosophila melanogaster receptor protein tyrosine kinase ALK splice
variant A mRNA, complete cds.
ACCESSION AF236106
VERSION AF236106.1 GI:7110574
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 8435)
Palmer, R.H., Scully, A. Jr., Edeen, P., Thomas, J., McKeown, M. Jr. and
Hunter, T.
Identification of a novel Drosophila melanogaster RTK: dALK
Unpublished
2 (bases 1 to 8435)
Palmer, R.H., Scully, A. Jr., Edeen, P., Thomas, J., McKeown, M. Jr. and
Hunter, T.
Direct Submission
Submitted (17-FEB-2000) UCMP, Umea University, Building 6L, Umea
University, Umea S-901 87, Sweden
Location/Qualifiers
1..8435
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2"
/map="53C-D"
378..5483
/codon_start=1
/product="receptor protein tyrosine kinase ALK splice
variant A"
/protein_id="AAF36990.1"
/db_xref="GI:7110575"
/translation="MPQLRVLVVLAISLLVDPSPFAQRPLIMNHTFSLLPMQSPHSF
NSWPSASGGKQLNRGLHEQLPLPQTLFVATQYEDYEAAPGSRDNKRRLAQMAG
RPGSGNGRRGMRMESLRKELMNPVSGGPGGSGGSGYSATSSITRIGLGGIG
SADRYGALVDLRQPSGMAATAGPLGGYSTRPPLVAGVDQDQPSRNSKNSISELY
KLKALNQADEPAGVTHGNEFGDQPSADPQLEIKDRIDTSPNANTYAPDTATP
SSYEYELGVKCNFETPCSWTNGNSDGFVITGTLSKLNKLLGLLPGPAANDSIDAN
GHFLYARVNPSSRLNLTSPFSTMEKCEFLVHMOSDMSHGLSRVSVVLLHTAES
WVPAEILGDNVQWTRVLRGRVSRDFRIVFVDPDLRVGQGHVALDNLRMVNCFP
EGTKSEKSTSQVKCTSSKVPVCIHLPICRIDTDCDEADEQSCDKIPYGRCDFE
EDMCKWRDCKTTLTWSRHFGSSPTDGTGPDGHTQHLQNTSGYMLVNMQHNN
SEKNSIIGFASNAIMYSKTNPPSPVHGNPDSPYRNSCVVRFTHOFGKNPSTLSV
VENKRENTITLWSTKNGSDWMAEYVLPNTSKYIYLFQFARMGMYISDAVDD
FSLSPCEGLNIPEDHLGGYVWDRQNLKSPYKDFEYTNILEYTCDDTRGIVPSQ
AQCEAAVREQNTKTVLRHVVEDQSSYKGMQKVPHEGHYTIILAKGASGLGSGV
GSSRGVAVAILHLKNEELYFLVGOGENACIKSMGLKACAGCTDHDLDLAQYSFR
SKDMVKNIYIENGAGGGGSGYVFLNQAKNEAPLVVAGGGGIGIGYIDEDFOH
GQAKPLQAPESQINGEPGKKTAGPGGWRAKEDQALSPTYGAALLQGGRGHSCY
VELADNGTSVRRHGGGFGGGGCGCTGGGGYAGDVYLTESNGEGSSYISPSRS
LRSEITHAGASSPGCAITIIIPAEIGCGCDYRCVADDEFKSRVCICPDGWSLKRNH
TACEIREAGKSSFOVLVSLMISLAVLFICIAALIFMLYNRYORKKSKRKLMLVE
QDLQTLRLNNIDDSNLNPNVNGCDILNHDVNSLPQVARDLSLOLVNALKGAF
GEVYMALYRHRDGDVEMGVAVKTLREDPKREKEEDFLKEAATMAKFNHNMVLLGV
CFDRQYIYVLELLAGDGLQKFLNENRTPERFSLTMKDLLEFALDVAKCYRMESK
RFTHRDIAARNCLSSKGPGRVVIADFGMSRDYISDYIRKGGKAMLPKRWMPPEAF
LDGIFTSKTDWNSFGILLWEVFLSGRSPYQGHNTQVMELVVRGRLGSGTECPVSIY
KVMADCNPTPEDRPTFTLLEHLTACTQDASIMNAPLPNLPILPTASERDVTIRPN
GEEFLAVPDYLVPLPPGNNPNSMAGSGYVPELROOISSCTPPAVTSPAAPHPRP
VENIAPTSGCWETSRLPNSKVEPPVVGSHDVPLAGGEAKLISLDTQPPTTIO
PPLSPASQLDGLTDLPSALTSLPNGNAKOSYANQVKNAEAKVINGVNGVTVMNG
DVSTGGSGADSPFTIOGAYERYKDNHNSHISC"
BASE COUNT 2522 a 1856 c 1961 g 2096 t
ORIGIN
Query Match 2.7%; Score 20; DB 5; Length 8435;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 tcggcaagaatccggcgagc 204
|||||
Db 2215 TCGCAAGAATCCGGCGAGT 2234
```

```

RESULT 4
AC004287/C
LOCUS
DEFINITION
Drosophila melanogaster DNA sequence (P1 DS02309 (D166)), complete
sequence.
ACCESSION
AC004287 AC003783 AC003310 AC003311 AC003784 AC003312 AC003313
VERSION
AC003314 AC003315 AC003316 AC003317
KEYWORDS
AC004287.1 GI:3769311
SOURCE
Drosophila melanogaster (Subclones in tet from p1 clone DS02309
(D166)) DNA.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 70841)
AUTHORS
Celiker, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,
Svirskas, R.R., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E.,
Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,
Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,
Snir, E., Twomey, B., Wan, K.H., Zhang, R., Zieran, L.L. and Rubin, G.M.
Sequencing of Drosophila chromosome 2R, region 53C10-53D3
Unpublished (1997)
2 (bases 1 to 70841)
AUTHORS
Celiker, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,
Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,
Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,
Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,
Zieran, L.L. and Kimmel, B.E.
Direct Submission
Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Oct 19, 1998 this sequence version replaced gi:3766105.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to bdg@fruitfly.berkeley.edu.
Library location: 5-25.
FEATURES
source
Location/Qualifiers
1..70841
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="53C10-53D3"
/clone="P1 DS02309 (D166)"
BASE COUNT 20303 a 15224 c 15085 g 20229 t
ORIGIN
Query Match 2.7%; Score 20; DB 4; Length 70841;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 tcggcaagaatccggcact 204
|||||
DB 5700 TCGGCAAGAATCCGGCAGT 48407
RESULT 5
AC017187
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
AC017187
LOCUS
DEFINITION
Drosophila melanogaster (Subclones in tet from p1 clone DS02309
(D166)) DNA.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 93588)
AUTHORS
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10209809 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* this sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..93588
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 26957 a 19517 c 19788 g 27326 t
ORIGIN
Query Match 2.7%; Score 20; DB 64; Length 93588;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 tcggcaagaatccggcact 204
|||||
DB 5700 TCGGCAAGAATCCGGCAGT 5719
RESULT 6
AC003806/C
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386047 section 37
of 52, complete sequence.
ACCESSION
AC003806 AE002787
VERSION
AE003806.2 GI:10727503
KEYWORDS
HTG.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 268219)
AUTHORS
Adams, M.D., Celiker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
Andrews-Pfannkuch, C., Baldwin, D., Balw, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,
Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cavley, S., Dahlke, C., Davenport, L.B., Davies, P., de
Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S.,
Fleischmann, W., Flosser, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kensler, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,

```

Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarri, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, J. D., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T. I., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstein, G. M., Weissbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, C., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M. and Venter, J. C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 268219)
Adams, M. D., Celniker, S. E., Gibbs, R. A., Rubin, G. M. and Venter, C. J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7302890.
Location/Qualifiers
i. 268219
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
<9709. >9957
/gene="CG16797"
/product="CT35933"
/db_xref="FLYBASE:FBan0016797"
/db_xref="FLYBASE:FBgn0034130"
/evidence=not_experimental
<9709. >9957
/gene="CG16797"
/map="53C2-53C3"
/db_xref="FLYBASE:FBan0016797"
/db_xref="FLYBASE:FBgn0034130"
/evidence=not_experimental
9709. 9957
/gene="CG16797"
/note="CG16797 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0016797"
/db_xref="FLYBASE:FBgn0034130"
/evidence=not_experimental
/protein_id="AAF57994.1"
/translation="MLELILNVVHLSLQISYEFEIYESIYSDGSDVPPTLNISG
LFDLNTVTTITADGLSVSGNMTSVFNAQPTDRIEYFA"
complement(join(27841. .27954,28010. .>28540))
/gene="CG15712"
/product="CT35937"
/db_xref="FLYBASE:FBan0015712"
/db_xref="FLYBASE:FBgn0034131"
/evidence=not_experimental
complement(<27841. .>28540)
/gene="CG15712"
/map="53C3-53C4"
/db_xref="FLYBASE:FBan0015712"
/db_xref="FLYBASE:FBgn0034131"
/evidence=not_experimental
complement(join(27841. .27954,28010. .28177))
/gene="CG15712"
/note="CG15712 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015712"
/db_xref="FLYBASE:FBgn0034131"
/evidence=not_experimental
/protein_id="AAF57993.1"
/db_xref="GI:7302921"

mrna
/translation="MLRPLNLTFLWDVHSLPRTRIYSLIAQSKOSTEPFAKLYKALR
DKFKELQVRLSRDLQAPDKKLKSOKSINVDEILQIVFEVSWGPKTS"
join(34139. .34418,34498. .36233)
/gene="CG4439"
/product="CT14432"
/db_xref="FLYBASE:FBan0004439"
/db_xref="FLYBASE:FBgn00034132"
<34139. .>36233
/gene="CG4439"
/map="53C3-53C4"
/db_xref="FLYBASE:FBan0004439"
/db_xref="FLYBASE:FBgn0034132"
join(34317. .34418,34498. .36000)
/gene="CG4439"
/note="CG4439 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004439"
/db_xref="FLYBASE:FBgn0034132"
/protein_id="AAF57992.1"
/db_xref="GI:7302920"
/translation="MTRSQMISSLRSTLGLFRCHKALIFARIPAFQOVCKSEDGSC
NTNLQGVVGLYAKDKGKGLKSPGKEFDDRVGKRTITELIKESGLNGELGVRGLYQ
NIDKEYWAVAVGLGEGAGFNAEVIDEGMENVRVCAVGARALQOQCTCHVDGM
EYPOAEGAAMAVRWYNNVKNKRIATPKLEYGSTDODANTRGLFKAESONLARR
LSDTPANQTPSIPQAQTVDALPCPGSVSEVRSDMDIETQNLNSFLWVAGSGEPIIL
LEVSYCGTSPEERPIMLGKGLTFNSGGLCHPKRGMDEYRGAVSGAACAATLAAA
ALSPLINSAVLPCENPSGMATKPGDVVTVLLNGTMRKIDISLAGTVLLADPLLVA
QYTFKPLVVEGVSGMSGIRKGASATGLWTNNSTLWKNFKAGALTGRLWRMPLW
RSTFKLVAPRASYDICTSGEGHASSCLAAAILVELVPCSDWVHLDTHCTGMLAKHVP
PYLLKOCMTGRPTRESIIQFLYQMAK"
complement(join(39980. .42020,42187. .>42430))
/gene="CG7813"
/product="CT23722"
/db_xref="FLYBASE:FBan0007813"
/db_xref="FLYBASE:FBgn0034133"
complement(<39980. .>42430)
/gene="CG7813"
/map="53C3-53C4"
/db_xref="FLYBASE:FBan0007813"
/db_xref="FLYBASE:FBgn0034133"
complement(join(39980. .42020,42187. .42350))
/gene="CG7813"
/note="CG7813 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007813"
/db_xref="FLYBASE:FBgn0034133"
/protein_id="AAF57991.1"
/db_xref="GI:7302919"
/translation="MWEKLVNCIKSNGNGVKNVSCQVVELADLMKQVPPNQMHKFK
MPAKHBEYKDRVKRYEPESMPTIDWEYIRQNVREEFVDWVKGYETKYDKLHVSFENRH
AIVDHRKRYFLVDEEKVVTKCISEYKAESDKRIQELTEKLEFKVAMRPSYSEMTMEEF
CFARPHLAPDFINKPTFWPTPEQMPGSDPDAALAHHEEPPEPKPKPEKPLPK
PSGEGKPPAAAKPEKPEPVVDTSQAELAKLIAKDLMAKAIVLNLSLKKKMSGLAK
VOKKADAAKAARSETASKSATPAATKSLDSITERESGNICNOTIIRSEEAENPE
VKARHTNLSLEADPCDQAEERAREIARTLRKQKQKAEWEKHKKKDCDCAKETID
PCKLMEETVCKPDKPCKAEEAVCPDPCQKEDDDICKPDKPKDEBEGSCEDEE
DPCKKPKCDGDYQFGDNEGSTKSKQDQVFINISECSKELAKESKSGSGKPK
EKASLSIPOGSDQKPVLMQITKESKQDKHKAIEGANEIGPVYTDPNQLAELLVQ
KKEKPVIIIEAKPEDDKPITIVPKLEISANPKGDQVQKQAEQKSPKDNKQVFTWA
SGAATLLTEATNTLEDLKRKEARLEALQAYTSAQROGALAEASKAVANANLAI
RSAEQTGEVSSRDREALEMAEKHAILAKLAGRAUKDAIARVLDKLKKK"
join(59206. .59270,60019. .60341,68374. .68461,68713. .68764,
70614. .70682,71542. .71649,71899. .72036,72108. .72240,
72359. .72715,73250. .73478,73729. .74065,74232. .74395,
74459. .74611,74794. .74902,75329. .75664)
/gene="Sema-2a"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT42076"
/db_xref="FLYBASE:FBan0004700"
/db_xref="FLYBASE:FBgn0011260"
<59206. .>75664
/gene="Sema-2a"

gene

```

/note="CG4700"
/map="53C4-53C6"
/db_xref="FLYBASE:FBan0004700"
join(59220..59270,59920..60341,68374..68461,68713..68764,
70614..70682,71542..71649,71899..72036,72108..72240,
72359..72715,73250..73478,73729..74065,74232..74395,
74459..74611,74794..74902,75329..75664)
/gene="Sema-2a"
/product="CT14922"
/db_xref="FLYBASE:FBan0004700"
join(60176..60341,68374..68461,68713..68764,70614..70682,
71542..71649,71899..72036,72108..72240,72359..72715,
73250..73478,73729..74065,74232..74395,74459..74611,
74794..74902,75329..75400)
/gene="Sema-2a"
/note="Sema-2a gene product [alt 1]"
/codon_start=1

Query Match      2.7%; Score 20; DB 5; Length 268219;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 tcggcaagaatccggcagc 204
|||||
Db 175024 TCGGCAAGATCGGGCAGT 175005

RESULT 7
LOCUS HSU45983 1944 bp DNA PRI 01-JUL-1997
DEFINITION Homo sapiens CCR8 chemokine receptor (CKBR8) gene, complete cds.
ACCESSION U45983
VERSION U45983.1 GI:2231165
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1944)
AUTHORS Tiffany,H.L., Lautens,L.L., Gao,J.-L., Pease,J., Locati,M.,
Combadere,C., Modi,W., Bonner,T.I. and Murphy,P.M.
TITLE Identification of a human monocyte and thymus receptor for the CC
chemokine I-309
JOURNAL J. Exp. Med. (1997) In press
REFERENCE 2 (bases 1 to 1944)
AUTHORS Bonner,T.I.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1996) Tom I. Bonner, Lab of Cell Biology, NIMH,
Bldg 36, Rm 3A-17, MSC 4090, Bethesda, MD 20892-4090, USA
REFERENCE 3 (bases 1 to 1944)
AUTHORS Bonner,T.I.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1997) Tom I. Bonner, Lab of Cell Biology, NIMH,
Bldg 36, Rm 3A-17, MSC 4090, Bethesda, MD 20892-4090, USA
REMARK Sequence update by submitter
COMMENT On Jul 1, 1997 this sequence version replaced gi:1245056.
FEATURES
source
1..1944
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p22-p22; between D3S1260 and D3S322"
<1..243
/note="based on comparison to GenBank Accession Number
AF005210"
evidence=not_experimental
gene 258..1325
/gene="CKBR8"
CDS 258..1325
/gene="CKBR8"
/note="GPR-CY6; G protein-coupled receptor"

/codon_start=1
/product="CCR8 chemokine receptor"
/protein_id="AAB61962.1"
/db_xref="GI:1245057"
/translation="MDYTLDSLVTVDYYPDISSPCDAELIQTNGKLLAVFYCL
LFVSLGNSLVILVVCVKLRSLTDLVLLNLSLSDLLFVFFPOTYLLDQWVFG
TMVKVSGFYIGFYSSMFFITLMSVDRLAVHVAVALKVRTIRMGTTCLAVWLT
AIMATIPLLVYQVASEDGLQCYSFYNQOTLKKWIKFTNFKMNLGLLIPFTIMFCY
IKILHLKRCQNHKTKAIRLVLIIVIASLLFLWPFNVVLFUTSLHSHMLDGCISQ
QLTYATHVTEIISFTHCCPNPVIYAFVGEKFKHLSEIFQKSCSIFNYLGRMPRES
CEKSSSQHSSRSSSDVYL"
547 a 422 c 410 g 565 t

BASE COUNT
ORIGIN
Query Match      2.6%; Score 19; DB 97; Length 1944;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 agaagtgatattgaatt 183
|||||
Db 1644 AGAAGTGGATATTGAAGTT 1662

RESULT 8
LOCUS HSU62556 2608 bp DNA PRI 02-AUG-1996
DEFINITION Human chemokine receptor-like protein (TER1) gene, complete cds.
ACCESSION U62556
VERSION U62556.1 GI:1468978
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2608)
AUTHORS Napolitano,M., Zingoni,A., Bernardini,G., Spinetti,G., Rocchi,M.
and Santoni,A.
TITLE Human chemokine receptor-like TER1 gene
JOURNAL J. Immunol. (1996) In press
REFERENCE 2 (bases 1 to 2608)
AUTHORS Napolitano,M., Zingoni,A., Bernardini,G., Spinetti,G., Rocchi,M.
and Santoni,A.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1996) Laboratory of Pathophysiology, Regina Elena
Cancer Institute, via delle Messi d'Oro 156, Rome 00158, Italy
FEATURES
source
1..2608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21"
/rpt_family="Alu"
repeat_region 1..142
gene 361..2200
/gene="TER1"
CDS 361..1428
/gene="TER1"
/note="Seven trans-membrane spanning chemokine-like
receptor"
/codon_start=1
/product="chemokine receptor-like protein"
/protein_id="AAB05542.1"
/db_xref="GI:1468979"
/translation="MDYTLDSLVTVDYYPDISSPCDAELIQTNGKLLAVFYCL
LFVSLGNSLVILVVCVKLRSLTDLVLLNLSLSDLLFVFFPOTYLLDQWVFG
TMVKVSGFYIGFYSSMFFITLMSVDRLAVHVAVALKVRTIRMGTTCLAVWLT
AIMATIPLLVYQVASEDGLQCYSFYNQOTLKKWIKFTNFKMNLGLLIPFTIMFCY
IKILHLKRCQNHKTKAIRLVLIIVIASLLFLWPFNVVLFUTSLHSHMLDGCISQ
QLTYATHVTEIISFTHCCPNPVIYAFVGEKFKHLSEIFQKSCSIFNYLGRMPRES
CEKSSSQHSSRSSSDVYL"
2143..2148
/gene="TER1"
polya_signal

```

```

repeat_region
1..142
/map_family="Alu"
361..2200
/gene="TER1"
361..1428
/gene="TER1"
/note="Seven trans-membrane spanning chemokine-like receptor"
/codon_start=1
/product="chemokine receptor-like protein"
/db_xref="AA805542.1"
/db_xref="GI:1468979"
/translation="MDYLDLSVTVTDYDDYDFSPSPDAELIQNGKLLLAIVFVCL
LFVSLGNSLVILVLVCKKRSITDVLNLAALLDLFVSPFPQTYLLDQWVFG
TVMKVSGSYIYIGFSYMEFFITMSVDRYLAVHVAIVALKVFTIRMGTTLCIAVWLT
AMATIPLLVYQVASEDGLOQYSFVNOOTLKWITFNFKMNLGLLPTFTFMFCY
IKTLHQLKQFNQYHNTKAIPLVLIVVIALFLFWPENVVLFTLSHMHLDGCSLSQ
OLYATHVTEIISFTCCCNVPYIAFVGEKFKHLSEIQKSCSQIFNYLGROMPRES
CEKSSSCQCHSSRSSSDVYIL"
2143..2148
/gene="TER1"
/note="putative"
polyA_signal

```

FEATURES	source	NOG 2V4, Canada
gene	1. .86172 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3" 14331. .15366 /gene="hnrNPAL" /pseudo	Location/Qualifiers
mRNA	complement(join(16138. .17996,20486. .20608)) /gene="CCR8" /note="transcript 1"	
mRNA	/product="CC chemokine receptor 8" complement(16138. .18030) /gene="CCR8" /note="transcript 2"	
gene	/product="CC chemokine receptor 8" complement(16138. .20608) /gene="CCR8"	
polyA_signal	complement(16138. .16143) /gene="CCR8" /note="putative"	
polyA_signal	complement(16177. .16182) /gene="CCR8" /note="putative"	
polyA_signal	complement(16183. .16189) /gene="CCR8" /note="putative"	

```
polyA_signal complement(16190. .16195)
/gene="CCR8"
/note="putative"
CDS complement(16915. .17982)
/gene="CCR8"
/codon_start=1
/product="CC chemokine receptor 8"
/protein_id="AAK08628.1"
/db_xref="GI:13027670"
/translation="MDYTLDSVTITVDYDDYDPDIFSSPCDGLIQTNGKLLAVFYCL
LFVSLGNSLVILVVCCKLRSLTDVLLNLASDLLFVSPFQTYILLDQWVFG
TMCKVSGFYIGYSMEFTILMSVDYRLAVHAYALKVIRTIRMGTTLCIAVLT
AIMATIPLLVFOVASDEGLQCYSFYNQOOLKWKIFTNFKMNLGLLIPFTIMFCY
IKILHLKRCQNHKTKAIRLVLIVIASLLFPVFNVLFTLSLHSHMLDGSISQ
OLTVAHTVEIISFTHCCVNPVIYAFVGEKPKHLSEIFQKSCQIFNVLGRQMPRES
CEKSSSCQOHSRSSVDYL"
join(68700. .68784,83882. .>86172)
/gene="CX3CR1"
/note="transcript 1"
/product="CX3C chemokine receptor 1"
/evidence=experimental
68700. .>86172
/gene="CX3CR1"
join(70361. .70441,83882. .>86172)
/note="transcript 2"
/product="CX3C chemokine receptor 1"
/evidence=experimental
83891. .84958
/gene="CX3CR1"
/codon_start=1
/product="CX3C chemokine receptor 1"
/protein_id="AAK08627.1"
/db_xref="GI:13027669"
/translation="MDQFPESVTENFEVDLAECYIGDIVVFGTVFLSIFYSVIFAI
GLVGNLVFALTSKPKSVTDIYLLNLASDLLFVATLPFWTHYLINERKGLHNCM
KFTTAFFFIGFGIFITVISIDRYLAIVLAANSNNRTVQHGVTISLGYWAAAILV
AAPQFMFTKQNECLGDYPEVLOEIWPLRNVTNPLGLLPLLLINSYCVFRIQTL
FSCNHHKKAAILKILVIVFLEFWPYNMIFLETLLKLYDFPSCDMRKDLRLALS
VTEVAFSHCCNPLIYAFAGEKFRRLYHLHYGKCLAVLCGRSVHVDSSSESQRSHR
GSVLSNNTYHTSDGALLLL"
BASE COUNT 23130 a 18485 c 17978 g 26342 t 237 others
ORIGIN

Query Match 2.6%; Score 19; DB 91; Length 86172;
Best Local Similarity 100.0%; Pred.No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 aagatggatattgaagtt 183
|||||
Db 15794 AGAAGTGGATATTGAAGTT 15812

RESULT 11
AY016370/c
LOCUS AY016370.1 86172 bp DNA PRI 21-FEB-2001
DEFINITION Homo sapiens hnrnpal pseudogene, complete sequence; and CC
chemokine receptor 8 (CCR8) and CX3C chemokine receptor 1 (CX3CR1)
genes, complete cds.
ACCESSION AY016370
VERSION AY016370.1 GI:13027668
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 86172)
DeVries, M.E., Xu, L., Kelvin, A.A. and Kelvin, D.J.
AUTHORS Identification and characterization of the CX3CR1 and CCR8
promoters
TITLE Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 86172)
```

Xu, L., DeVries, M.E., Kelvin, A.A. and Kelvin, D.J.
Direct Submission
Submitted (05-DEC-2000) Laboratory of Molecular Inflammation and
Immunology, Robarts Research Institute, 1400 Western Rd, London, ON
N6G 2V4, Canada
Location/Qualifiers
1. 86172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
14331. .15366
/gene="hnrnpal"
/pseudogene
complement(join(16138. .17996,20486. .20608))
/note="transcript 1"
/product="CC chemokine receptor 8"
complement(16138. .18030)
/gene="CCR8"
/note="transcript 2"
/product="CC chemokine receptor 8"
complement(16138. .20608)
/gene="CCR8"
complement(16138. .16143)
/gene="CCR8"
/note="putative"
complement(16177. .16182)
/gene="CCR8"
/note="putative"
complement(16183. .16189)
/gene="CCR8"
/note="putative"
complement(16190. .16195)
/gene="CCR8"
/note="putative"
complement(16915. .17982)
/gene="CCR8"
/codon_start=1
/product="CC chemokine receptor 8"
/protein_id="AAK08628.1"
/db_xref="GI:13027670"
/translation="MDYTLDSVTITVDYDDYDPDIFSSPCDGLIQTNGKLLAVFYCL
LFVSLGNSLVILVVCCKLRSLTDVLLNLASDLLFVSPFQTYILLDQWVFG
TMCKVSGFYIGYSMEFTILMSVDYRLAVHAYALKVIRTIRMGTTLCIAVLT
AIMATIPLLVFOVASDEGLQCYSFYNQOOLKWKIFTNFKMNLGLLIPFTIMFCY
IKILHLKRCQNHKTKAIRLVLIVIASLLFPVFNVLFTLSLHSHMLDGSISQ
OLTVAHTVEIISFTHCCVNPVIYAFVGEKPKHLSEIFQKSCQIFNVLGRQMPRES
CEKSSSCQOHSRSSVDYL"
join(68700. .68784,83882. .>86172)
/gene="CX3CR1"
/note="transcript 1"
/product="CX3C chemokine receptor 1"
/evidence=experimental
68700. .>86172
/gene="CX3CR1"
join(70361. .70441,83882. .>86172)
/note="transcript 2"
/product="CX3C chemokine receptor 1"
/evidence=experimental
83891. .84958
/gene="CX3CR1"
/codon_start=1
/product="CX3C chemokine receptor 1"
/protein_id="AAK08627.1"
/db_xref="GI:13027669"
/translation="MDQFPESVTENFEVDLAECYIGDIVVFGTVFLSIFYSVIFAI
GLVGNLVFALTSKPKSVTDIYLLNLASDLLFVATLPFWTHYLINERKGLHNCM
KFTTAFFFIGFGIFITVISIDRYLAIVLAANSNNRTVQHGVTISLGYWAAAILV
AAPQFMFTKQNECLGDYPEVLOEIWPLRNVTNPLGLLPLLLINSYCVFRIQTL
FSCNHHKKAAILKILVIVFLEFWPYNMIFLETLLKLYDFPSCDMRKDLRLALS
VTEVAFSHCCNPLIYAFAGEKFRRLYHLHYGKCLAVLCGRSVHVDSSSESQRSHR
GSVLSNNTYHTSDGALLLL"

BASE COUNT 23130 a 18485 c 17978 g 26342 t 237 others.

Query Match 2.6%; Score 19; DB 91; Length 86172;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 agaagtgtattgaagt 183
|||||
Db 16594 AGAAGTGGATATTGAAGTT 16576

RESULT 12

AL136961/c
LOCUS AL136961 98861 bp DNA PRI 11-APR-2001
DEFINITION Human DNA sequence from clone RP11-178C10 on chromosome 13,
complete sequence.
ACCESSION AL136961
VERSION AL136961.19 GI:13620283
KEYWORDS HTP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (Bases 1 to 98861)
AUTHORS Blakey, S.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clones@esanger.ac.uk

COMMENT

On Apr 12, 2001 this sequence version replaced gi:13276997.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; SW.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
RP11-178C10 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-178C10. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-178C10 is at 98861 in this
sequence. The true right end of clone RP11-461N23 is at 100 in this
sequence.

FEATURES

Location/Qualifiers
source 1..98861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-178C10"
/clone.lib="RPC1-11.1"
repeat_region 161..463
/note="L1MB8 repeat: matches 5836..6160 of consensus"

repeat_region 656..723
/note="L2 repeat: matches 2592..2659 of consensus"
repeat_region 2385..2420
/note="MIR repeat: matches 111..146 of consensus"
repeat_region 4555..4646
/note="L2 repeat: matches 1031..1120 of consensus"
repeat_region 4989..5306
/note="MER7A repeat: matches 2..345 of consensus"
repeat_region 5859..5987
/note="L2 repeat: matches 2574..2708 of consensus"
repeat_region 6215..6548
/note="Charliel repeat: matches 2292..2629 of consensus"
repeat_region 6557..6635
/note="L2 repeat: matches 2128..2206 of consensus"
repeat_region 6730..7233
/note="L2 repeat: matches 1330..1889 of consensus"
repeat_region 8301..8499
/note="MIR repeat: matches 17..219 of consensus"
repeat_region 8893..8994
/note="L2 repeat: matches 1551..1656 of consensus"
repeat_region 9079..9126
/note="24 copies 2 mer at 75% conserved"
repeat_region 9554..9754
/note="MER20 repeat: matches 33..218 of consensus"
repeat_region 9816..9979
/note="FAM repeat: matches 2..165 of consensus"
repeat_region 10402..10703
/note="L1M1 repeat: matches 5264..5577 of consensus"
repeat_region 12181..12626
/note="L1 repeat: matches 3474..3949 of consensus"
repeat_region 12654..13012
/note="L1 repeat: matches 5571..5927 of consensus"
repeat_region 13491..14076
/note="L1 repeat: matches 2226..2881 of consensus"
repeat_region 16310..17185
/note="L1M4 repeat: matches 6237..7151 of consensus"
repeat_region 17207..17622
/note="MER47B repeat: matches 2..413 of consensus"
repeat_region 17628..18380
/note="L1M4 repeat: matches 7197..7976 of consensus"
repeat_region 20797..21100
/note="AluY repeat: matches 1..297 of consensus"
repeat_region 22146..22814
/note="L2 repeat: matches 2574..2643 of consensus"
repeat_region 23758..23981
/note="AluJb repeat: matches 85..310 of consensus"
repeat_region 25108..25212
/note="L2 repeat: matches 2570..2688 of consensus"
repeat_region 25740..25795
/note="MIR repeat: matches 196..251 of consensus"
repeat_region 25982..26012
/note="Tigger4(Zombi) repeat: matches 2701..2731 of
consensus"
repeat_region 26005..26208
/note="Tigger4(Zombi) repeat: matches 1772..1970 of
consensus"
repeat_region 26209..26513
/note="AluX repeat: matches 1..305 of consensus"
repeat_region 26514..28201
/note="Tigger4(Zombi) repeat: matches 1..1772 of
consensus"
repeat_region 32616..32843
/note="6 copies 38 mer 86% conserved"
repeat_region 35333..35362
/note="15 copies 2 mer ac 100% conserved"
repeat_region 36220..36303
/note="2 copies 42 mer 97% conserved"
repeat_region 36967..37274
/note="AluSq repeat: matches 1..313 of consensus"
repeat_region 39481..39767
/note="AluX repeat: matches 1..297 of consensus"
repeat_region 41084..41111
/note="14 copies 2 mer aa 89% conserved"

repeat_region 4119..41174 /note="THE1C repeat: matches 1. .58 of consensus"
repeat_region 41175..41482 /note="AluSp repeat: matches 1. .309 of consensus"
repeat_region 41483..41785 /note="THE1C repeat: matches 58. .344 of consensus"
repeat_region 41791..41842 /note="L2 repeat: matches 1215. .1266 of consensus"
repeat_region 42051..42573 /note="MER74A repeat: matches 2. .555 of consensus"
repeat_region 42857..43014 /note="L2 repeat: matches 2553. .2733 of consensus"
repeat_region 43250..43279 /note="L5 copies 2 mer ca 100% conserved"
repeat_region 44262..44708 /note="MLT1C repeat: matches 1. .466 of consensus"
repeat_region 44860..45175 /note="MLT2FB repeat: matches 1. .328 of consensus"
repeat_region 45204..45235 /note="L16 copies 2 mer tg 87% conserved"
repeat_region 45389..45559 /note="MER58 repeat: matches 4. .175 of consensus"
repeat_region 45635..45788 /note="MER5A repeat: matches 9. .184 of consensus"
repeat_region 45963..46162 /note="MIR repeat: matches 25. .250 of consensus"
repeat_region 46281..46837 /note="MER67D repeat: matches 1. .510 of consensus"
repeat_region 47886..51275 /note="L1PA6 repeat: matches 2744. .6141 of consensus"
repeat_region 51364..51523 /note="MER5A repeat: matches 1. .189 of consensus"
repeat_region 51736..51791 /note="MIR repeat: matches 118. .178 of consensus"
repeat_region 51848..52167 /note="MSTB repeat: matches 86. .422 of consensus"
repeat_region 53138..53915 /note="MER21B repeat: matches 41. .789 of consensus"
repeat_region 54321..54629 /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 55247..55308 /note="MER69A repeat: matches 24. .87 of consensus"
repeat_region 55309..55622 /note="AluSg repeat: matches 1. .313 of consensus"
repeat_region 55623..55715 /note="MER59A repeat: matches 87. .176 of consensus"
repeat_region 57171..57271 /note="MLT1F repeat: matches 437. .541 of consensus"
repeat_region 57462..57560 /note="MLT1G repeat: matches 169. .264 of consensus"
repeat_region 57561..58049 /note="MER74A repeat: matches 1. .519 of consensus"
repeat_region 58063..58287 /note="AluSg/x repeat: matches 163. .300 of consensus"
repeat_region 58336..58475 /note="MLT1G repeat: matches 76. .205 of consensus"
repeat_region 58496..58809 /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region 58818..59182 /note="L1ME3A repeat: matches 5789. .6157 of consensus"
repeat_region 59198..59336 /note="MER74A repeat: matches 1. .139 of consensus"
repeat_region 59337..59435 /note="MLT1G repeat: matches 169. .264 of consensus"
repeat_region 59626..59726 /note="MLT1F repeat: matches 437. .541 of consensus"
repeat_region 59804..60249 /note="L1ME3A repeat: matches 5301. .5782 of consensus"
repeat_region 61972..62009

Query Match 2.6%, Score 19; DB 89; Length 98861;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 aagaatggatattgaagtt 183
|||||
Db 26530 AGAAGTGGATATTGAAGTT 26512

RESULT 13
AC084179
LOCUS
DEFINITION
AC084179
AC084179.2 GI:12229370
HTG: HTGS_PHASE1; HTGS_DRAFT.
human
Homo sapiens

ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147009)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 5, clone RP11-148P16
Unpublished
2 (bases 1 to 147009)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeBellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnaz,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Tsalas,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

AC084179 147009 bp DNA HTG 15-JAN-2001
Homo sapiens chromosome 5 clone RP11-148P16 map 5, WORKING DRAFT
SEQUENCE, 39 unordered pieces.
AC084179
HTG: HTGS_PHASE1; HTGS_DRAFT.
human
Homo sapiens

1 (bases 1 to 147009)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 5, clone RP11-148P16
Unpublished
2 (bases 1 to 147009)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeBellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnaz,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Tsalas,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

Direct Submission
Submitted (14-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 15, 2001 this sequence version replaced gi:10801407.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11428
Center clone name: 148_P_16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132233 bases at least Q40
Consensus quality: 139576 bases at least Q30
Consensus quality: 142136 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 143209; sum-of-contigs
Quality coverage: 3.0 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11428
Center clone name: 148_P_16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132233 bases at least Q40
Consensus quality: 139576 bases at least Q30
Consensus quality: 142136 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 143209; sum-of-contigs
Quality coverage: 3.0 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is

1	1353:	contig of 1353 bp in length
1354	1453:	gap of 100 bp
1454	2001:	contig of 548 bp in length
2002	2101:	gap of 100 bp
2102	3124:	contig of 1023 bp in length
3125	3224:	gap of 100 bp
3225	4546:	contig of 1322 bp in length
4547	4646:	gap of 100 bp
4647	6015:	contig of 1369 bp in length
6016	6115:	gap of 100 bp
6116	7299:	contig of 1184 bp in length
7300	7399:	gap of 100 bp
7400	8836:	contig of 1437 bp in length
8837	8936:	gap of 100 bp
8937	10282:	contig of 1346 bp in length
10283	10382:	gap of 100 bp
10383	11352:	contig of 970 bp in length
11353	11452:	gap of 100 bp
11453	12856:	contig of 1404 bp in length
12857	12956:	gap of 100 bp
12957	14187:	contig of 1231 bp in length
14188	14287:	gap of 100 bp
14288	15977:	contig of 1690 bp in length
15978	16077:	gap of 100 bp
16078	17844:	contig of 1767 bp in length
17845	17944:	gap of 100 bp
17945	19861:	contig of 1917 bp in length
19862	19961:	gap of 100 bp
19962	22368:	contig of 2407 bp in length
22369	22466:	gap of 100 bp
22469	24485:	contig of 2017 bp in length
24486	24585:	gap of 100 bp
24586	26722:	contig of 2137 bp in length
26723	26822:	gap of 100 bp
26823	28983:	contig of 2161 bp in length
28984	29083:	gap of 100 bp
29084	31858:	contig of 2775 bp in length
31859	31958:	gap of 100 bp
31959	33783:	contig of 1825 bp in length
33784	33883:	gap of 100 bp
33884	36259:	contig of 2376 bp in length
36260	36359:	gap of 100 bp
36360	39396:	contig of 3037 bp in length
39397	39496:	gap of 100 bp
39497	43329:	contig of 3833 bp in length
43330	43429:	gap of 100 bp
43430	46057:	contig of 2628 bp in length
46058	46157:	gap of 100 bp
46158	50066:	contig of 3909 bp in length
50067	50166:	gap of 100 bp
50167	53727:	contig of 3561 bp in length
53728	53827:	gap of 100 bp
53828	57631:	contig of 3804 bp in length
57632	57731:	gap of 100 bp
57732	61304:	contig of 3573 bp in length
61305	61404:	gap of 100 bp
61405	63868:	contig of 2464 bp in length
63869	63968:	gap of 100 bp
63969	68393:	contig of 4425 bp in length
68394	68493:	gap of 100 bp
68494	72466:	contig of 3373 bp in length
72467	72566:	gap of 100 bp
72567	77027:	contig of 4461 bp in length
77028	77127:	gap of 100 bp
77128	81524:	contig of 4397 bp in length
81525	81624:	gap of 100 bp
81625	88873:	contig of 7249 bp in length
88874	88973:	gap of 100 bp

AC016189

LOCUS AC016189 151174 bp DNA HTG 13-JUL-2000
 DEFINITION Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
 AC016189
 AC016189.3 GI:9129134
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 151174)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-23M9
 Unpublished
 2 (bases 1 to 151174)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckghalter,B.,
 Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McCurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi.6939915.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4399
 Center clone name: 23_M_9

* NOTE: This record contains 152 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 851: contig of 851 bp in length
 852 951: gap of 100 bp
 952 1804: contig of 853 bp in length
 1805 1904: gap of 100 bp
 1905 2719: contig of 815 bp in length
 2720 2819: gap of 100 bp
 2820 3668: contig of 849 bp in length
 3669 3768: gap of 100 bp
 3769 4657: contig of 889 bp in length
 4658 4757: gap of 100 bp
 4758 5600: contig of 843 bp in length
 5601 5700: gap of 100 bp
 5701 6532: contig of 832 bp in length
 6533 6632: gap of 100 bp
 6633 7481: contig of 849 bp in length
 7482 7581: gap of 100 bp
 7582 8447: contig of 866 bp in length

8448 8547: gap of 100 bp
 8548 9383: contig of 836 bp in length
 9384 9483: gap of 100 bp
 9484 10342: contig of 859 bp in length
 10343 10442: gap of 100 bp
 10443 11323: contig of 881 bp in length
 11324 11423: gap of 100 bp
 11424 12254: contig of 831 bp in length
 12255 12354: gap of 100 bp
 12355 13288: contig of 934 bp in length
 13289 13388: gap of 100 bp
 13389 14235: contig of 847 bp in length
 14236 14335: gap of 100 bp
 14336 15172: contig of 837 bp in length
 15173 15272: gap of 100 bp
 15273 16130: contig of 858 bp in length
 16131 16230: gap of 100 bp
 16231 17073: contig of 843 bp in length
 17074 17173: gap of 100 bp
 17174 18003: contig of 830 bp in length
 18004 18103: gap of 100 bp
 18104 18939: contig of 836 bp in length
 18940 19039: gap of 100 bp
 19040 19885: contig of 846 bp in length
 19886 19985: gap of 100 bp
 19986 20834: contig of 849 bp in length
 20835 20934: gap of 100 bp
 20935 21761: contig of 827 bp in length
 21762 21861: gap of 100 bp
 21862 22743: contig of 882 bp in length
 22744 22843: gap of 100 bp
 22844 23708: contig of 865 bp in length
 23709 23808: gap of 100 bp
 23809 24676: contig of 868 bp in length
 24677 24776: gap of 100 bp
 24777 25602: contig of 826 bp in length
 25603 25702: gap of 100 bp
 25703 26560: contig of 858 bp in length
 26561 26660: gap of 100 bp
 26661 27504: contig of 844 bp in length
 27505 27604: gap of 100 bp
 27605 28449: contig of 845 bp in length
 28450 28549: gap of 100 bp
 28550 29398: contig of 849 bp in length
 29399 29498: gap of 100 bp
 29499 30357: contig of 859 bp in length
 30358 30457: gap of 100 bp
 30458 31323: contig of 866 bp in length
 31324 31423: gap of 100 bp
 31424 32300: contig of 877 bp in length
 32301 32400: gap of 100 bp
 32401 33254: contig of 854 bp in length
 33255 33354: gap of 100 bp
 33355 34218: contig of 864 bp in length
 34219 34318: gap of 100 bp
 34319 35169: contig of 851 bp in length
 35170 35269: gap of 100 bp
 35270 36125: contig of 856 bp in length
 36126 36225: gap of 100 bp
 36226 37061: contig of 836 bp in length
 37062 37161: gap of 100 bp
 37162 38003: contig of 842 bp in length
 38004 38103: gap of 100 bp
 38104 38945: contig of 842 bp in length
 38946 39045: gap of 100 bp
 39046 39882: contig of 837 bp in length
 39883 39982: gap of 100 bp
 39983 40833: contig of 851 bp in length
 40834 40933: gap of 100 bp
 40934 41778: contig of 846 bp in length
 41780 41879: gap of 100 bp
 41880 42722: contig of 843 bp in length
 42723 42822: gap of 100 bp

```

* 42823 43678: contig of 856 bp in length
* 43679 43778: gap of 100 bp
* 43779 44634: contig of 856 bp in length
* 44635 44734: gap of 100 bp
* 44735 45587: contig of 853 bp in length
* 45588 45687: gap of 100 bp
* 45688 46542: contig of 855 bp in length
* 46543 46642: gap of 100 bp
* 46643 47488: contig of 846 bp in length
* 47489 47588: gap of 100 bp
* 47589 48448: contig of 860 bp in length
* 48449 48548: gap of 100 bp
* 48549 49450: contig of 902 bp in length
* 49451 49550: gap of 100 bp
* 49551 50383: contig of 833 bp in length
* 50384 50483: gap of 100 bp
* 50484 51353: contig of 870 bp in length
* 51354 51453: gap of 100 bp
* 51454 52308: contig of 853 bp in length
* 52309 52408: gap of 100 bp
* 52409 53252: contig of 846 bp in length
* 53253 53352: gap of 100 bp
* 53353 54224: contig of 872 bp in length
* 54225 54324: gap of 100 bp
* 54325 55162: contig of 838 bp in length
* 55163 55262: gap of 100 bp
* 55263 56086: contig of 824 bp in length
* 56087 56186: gap of 100 bp
* 56187 57026: contig of 840 bp in length
* 57027 57126: gap of 100 bp
* 57127 58025: contig of 899 bp in length
* 58026 58125: gap of 100 bp
* 58126 59004: contig of 879 bp in length
* 59005 59104: gap of 100 bp
* 59105 59979: contig of 875 bp in length
* 59980 60078: gap of 100 bp
* 60080 60920: contig of 841 bp in length
* 60921 61020: gap of 100 bp
* 61021 61885: contig of 865 bp in length
* 61886 61985: gap of 100 bp
* 61986 62792: contig of 807 bp in length
* 62793 62892: gap of 100 bp
* 62893 63744: contig of 852 bp in length
* 63745 63844: gap of 100 bp
* 63845 64693: contig of 849 bp in length
* 64694 64793: gap of 100 bp
* 64794 65645: contig of 852 bp in length
* 65646 65745: gap of 100 bp
* 65746 66624: contig of 879 bp in length
* 66625 66724: gap of 100 bp
* 66725 67550: contig of 826 bp in length
* 67551 67650: gap of 100 bp
* 67651 68502: contig of 852 bp in length
* 68503 68602: gap of 100 bp
* 68603 69435: contig of 833 bp in length

```

```

Query Match      2.6%; Score 19; DB 64; Length 151174;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 165 aagatggatattgaagtt 183
|||||
Db 119794 AGAAGTGGATATTGAAGTT 119812

```

```

RESULT 15
AC026309
LOCUS AC026309 159852 bp DNA HTG 17-FEB-2001
DEFINITION Homo sapiens chromosome 3 clone RP11-256B12, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC026309
VERSION AC026309.16 GI:12831263
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

SOURCE ORGANISM

human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 159852)

REFERENCE AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brleva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louliseh,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaikhe,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
and Gibbs,R.

TITLE

Direct Submission

Unpublished

2 (bases 1 to 159852)

Worley,K.C.

TITLE

Direct Submission

Submitted (22-MAR-2000)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Feb 15, 2001 this sequence version replaced gi:9930647.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HAPA

Center clone name: RP11-256B12

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-primer Bodipy: 70% of reads

Chemistry: Dye-terminator Big Dye: 30% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 146110 bases at least Q40

Consensus quality: 151838 bases at least Q30

Consensus quality: 154679 bases at least Q20

Estimated insert size: 154344; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 3.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 consists of 21 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 22773: contig of 22773 bp in length
 22774 22873: gap of unknown length
 22874 41510: contig of 18637 bp in length
 41511 41610: gap of unknown length
 41611 54121: contig of 12511 bp in length
 54122 54221: gap of unknown length
 54222 65464: contig of 11243 bp in length
 65465 65565: gap of unknown length
 65566 74587: contig of 9023 bp in length
 74588 74688: gap of unknown length
 74689 82976: contig of 8289 bp in length
 82977 83077: gap of unknown length
 83078 92749: contig of 9673 bp in length
 92750 92849: gap of unknown length
 92850 101272: contig of 8423 bp in length
 101273 101372: gap of unknown length
 101373 109844: contig of 8472 bp in length
 109845 109944: gap of unknown length
 109945 117434: contig of 7489 bp in length
 117435 117534: gap of unknown length
 117535 121707: contig of 4174 bp in length
 121708 121807: gap of unknown length
 121808 127457: contig of 5649 bp in length
 127458 127557: gap of unknown length
 127558 132508: contig of 4951 bp in length
 132509 132607: gap of unknown length
 132608 137494: contig of 4887 bp in length
 137495 137594: gap of unknown length
 137595 141045: contig of 3451 bp in length
 141046 141145: gap of unknown length
 141146 144615: contig of 3470 bp in length
 144616 144715: gap of unknown length
 144716 147015: contig of 2300 bp in length
 147016 147115: gap of unknown length
 147116 149623: contig of 2508 bp in length
 149624 149723: gap of unknown length
 149724 152867: contig of 3144 bp in length
 152868 152967: gap of unknown length
 152968 156902: contig of 3935 bp in length
 156903 157002: gap of unknown length
 157003 159852: contig of 2850 bp in length.

FEATURES
 Source

1..159852
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /Chromosome="3"
 /clone="RP11-256B12"

BASE COUNT 45632 a 34294 c 33182 g 44726 t 2018 others
 ORIGIN

Query Match 2.6%; Score 19; DB 70; Length 159852;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 agaagtcgatatgaagtt 183
 |||||
 Db 77555 AGAAGTGGATTTGAAGTT 77573

Search completed: August 23, 2001, 15:04:24
 Job time: 10211 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:17:22 ; Search time 157.23 Seconds
(without alignments)
2971.181 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgctccg 744

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	2.6	609	21	Fusarium venenatum
2	19	2.6	1944	20	Human CC chemokine
3	17	2.3	250	15	Human genome fragm
4	17	2.3	525	20	Human CSR protein
5	17	2.3	824	21	Sheep beta-lactogl
6	17	2.3	824	22	Sheep beta-lactogl
7	17	2.3	3564	20	Human Y chromosome
8	17	2.3	3760	20	Borrelia burgdorfe
9	17	2.3	6763	18	Beta-lactoglobulin
10	17	2.3	7494	18	Staphylococcus aur
11	17	2.3	10807	16	Sheep beta-lactogl

C 12	17	2.3	10807	18	AAAT79725	Ovine beta-lactogl
C 13	17	2.3	12061	22	AAF28652	Partial DNA sequen
C 14	17	2.3	15144	18	AAAT77115	Mouse bone morphog
C 15	16	2.2	33	21	AAA53077	Rat genomic DNA ve
C 16	16	2.2	147	14	AAQ38649	Encodes antifungal
C 17	16	2.2	147	15	AAQ70127	Antimicrobial Ct-A
C 18	16	2.2	303	21	AAQ08635	Human secreted pro
C 19	16	2.2	343	20	AAV87787	EST clone ET13. H
C 20	16	2.2	384	21	AAQ00698	Human secreted pro
C 21	16	2.2	397	20	AAQ89901	MS2 bacteriophage
C 22	16	2.2	478	21	AAQ54507	Arabidopsis thalia
C 23	16	2.2	479	21	AAQ53279	Arabidopsis thalia
C 24	16	2.2	594	19	AAQ98649	DNA encoding a S.
C 25	16	2.2	634	19	AAV52656	Human native hepat
C 26	16	2.2	653	21	AAV13097	Aspergillus oryzae
C 27	16	2.2	746	21	AAQ12605	Aspergillus oryzae
C 28	16	2.2	770	21	AAQ26389	Human secreted pro
C 29	16	2.2	1100	20	AAQ08328	Thermus thermophil
C 30	16	2.2	1130	21	AAQ46203	Arabidopsis thalia
C 31	16	2.2	1133	21	AAQ39316	Arabidopsis thalia
C 32	16	2.2	1200	22	AAA54494	Botulin toxin hea
C 33	16	2.2	1365	22	AAQ88109	Human FLEXHT-40 nu
C 34	16	2.2	1386	22	AAQ81363	Quorum sensing con
C 35	16	2.2	1407	19	AAV26297	Phosphate starvati
C 36	16	2.2	1431	21	AAA59945	Polynucleotide seq
C 37	16	2.2	1587	21	AAA65734	Streptococcus pneu
C 38	16	2.2	1722	18	AAV74718	Staphylococcus aur
C 39	16	2.2	1777	20	AAV84538	Human secreted pro
C 40	16	2.2	2192	19	AAQ26284	S. pneumoniae deri
C 41	16	2.2	2528	21	AAQ65738	Streptococcus pneu
C 42	16	2.2	2565	16	AAQ86895	A. nidulans fksA g
C 43	16	2.2	2696	20	AAQ78223	Melon MEETR1 cDNA
C 44	16	2.2	3094	16	AAQ91398	Porphyrromonas ging
C 45	16	2.2	3120	21	AAQ65730	Streptococcus pneu

ALIGNMENTS

RESULT 1
AAQ08911
ID AAF08911 standard; cDNA; 609 BP.
XX AAF08911;
AC AAF08911;
DT 13-MAR-2001 (first entry)
XX Fusarium venenatum EST SEQ ID NO:1434.
DE Fusarium venenatum
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX Fusarium venenatum.
XX WO200056762-A2.
XX 28-SEP-2000.
XX 22-MAR-2000; 2000WO-US07781.
XX 22-MAR-1999; 99US-0273623.
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
WPI; 2000-594572/56.
XX Monitoring differential expression of genes in filamentous fungal cells
PT

PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 PS Claim 86; Page 924; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 609 BP; 139 A; 197 C; 147 G; 120 T; 6 other;

Query Match 2.6%; Score 19; DB 21; Length 609;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 485 tcataactgggtcaaggt 503
 |||||
 DB 235 tcataactgggtcaaggt 253

RESULT 2

AA024386
 ID AAX24386 standard; DNA; 1944 BP.

AC AAX24386;

DT 07-JUN-1999 (first entry)

XX Human CC chemokine receptor CCR8 DNA.

XX Chemokine receptor; CCR8; human; G protein coupled receptor; HIV;
 KW infection; therapy; immunomodulator; chemotaxis; apoptosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 258..1325
 FT /*tag= a

XX W09906561-A2.

XX 11-FEB-1999.

XX 29-JUL-1998; 98WO-US15730.

XX 29-JUL-1997; 97US-0054094.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Alkhatib G, Bazan H, Berger EA, Bonner TI, Lautens L;
 PI Murphy PM, Tiffany HL;

XX WPI: 1899-153791/13.
 DR P-PSDB; AAW97868.
 XX
 PT New isolated chemokine receptor CCR8 - used to develop agents for
 PT modulating immune responses or agents for the prevention or
 PT treatment of HIV infection
 XX
 PS Claim 13; Fig 10A; 81pp; English.
 XX
 CC This is the nucleotide sequence of a DNA clone encoding a novel
 CC human CC chemokine receptor (see AAW97868) that has been designated
 CC CCR8. CCR8 is a G protein coupled receptor that is a necessary
 CC cofactor for infection by HIV isolates. It plays an essential role
 CC in the membrane fusion step of infection, and interacts with
 CC envelopes from all 3 classes of HIV-1. Degenerate PCR primers (see
 CC AAX24387-88) based on the predicted 3rd and 7th transmembrane domains
 CC of CXCR2 and an orphan receptor named 9-6 were used to amplify a
 CC human genomic sequence named CV6. A 1953 bp fragment of a
 CC genomic clone containing CV6 was isolated and sequenced. It
 CC included an open reading frame for a polypeptide termed CV6/TER1
 CC ORF CC chemokine receptor 6 or CCR8. The gene was mapped to
 CC human chromosome 3p22-p23 by FISH. CCR8 is expressed constitutively
 CC in monocytes and thymus, and may regulate monocyte chemotaxis and
 CC thymic cell line apoptosis. The susceptibility of target cells to
 CC HIV infection depends on cell surface expression of human CD4 and
 CC CCR8. Establishment of stable, non-human cell lines and transgenic
 CC mammals having cells that coexpress CD4 and CCR8 provides valuable
 CC tools for research on HIV infection. Antibodies which bind to
 CC CCR8, CCR8 variants, and CCR8-binding agents capable of blocking
 CC membrane fusion between HIV and target cells represent potential
 CC anti-HIV therapeutics.

XX Sequence 1944 BP; 547 A; 422 C; 410 G; 565 T; 0 other;

Query Match 2.6%; Score 19; DB 20; Length 1944;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 agaagtggatattgaagtt 183
 |||||

DB 1644 agaagtggatattgaagtt 1662

RESULT 3

AA076996

ID AA076996 standard; DNA; 250 BP.

XX AC AA076996;

XX DT 23-SEP-1994 (first entry)

XX DE Human genome fragment.

XX KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
 KW detection; homology; human; adrenal tissue; ds.

XX OS Homo sapiens.

XX PN W09401548-A.

XX PD 20-JAN-1994.

XX PF 13-JUL-1993; 93WO-GB01467.

XX PR 13-JUL-1992; 92GB-0014857.

XX PA (MEDIC) MEDICAL RES COUNCIL.

XX PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
 XX Sibson DR, Starkey M;

DR WPI; 1994-035056/04.
 XX New nucleic acid fragment encoding gene products - can be used
 PT for genetic analysis and mapping
 XX
 XX Claim 1; Page 309; 616pp; English.
 XX
 CC Human nucleic acid fragments, isolated from brain adrenal tissue,
 CC the placenta or bone marrow comprise any of: (A) a sequence
 CC selected from (AAQ76401-Q77613), (B) an allelic variation of a
 CC sequence as described in (A), or (C) a sequence complementary
 CC to (A) or (B).
 XX
 SQ Sequence 250 BP; 71 A; 63 C; 63 G; 53 T; 0 other;

Query Match 2.3%; Score 17; DB 15; Length 250;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 tccagtcacacatt 222
 Db 59 tccagtcacacatt 75
 |||||

RESULT 4
 AAX27861
 ID AAX27861 standard; DNA; 525 BP.
 XX
 AC AAX27861;
 XX
 DT 02-JUN-1999 (first entry)
 XX
 DE Human CSR protein coding sequence fragment.
 XX
 KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
 KW scavenger receptor protein; intracellular stress; arteriosclerosis;
 KW diabetic circulatory obstruction; microbial infection; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9909159-A1.
 XX
 PD 25-FEB-1999.
 XX
 PF 12-AUG-1998; 98WO-JP03602.
 XX
 PR 30-JUL-1998; 98JP-0230121.
 PR 13-AUG-1997; 97JP-0233396.
 XX
 PA (NISR) JAPAN TOBACCO INC.
 XX
 PI Nakamura Y, Tokino T;
 XX
 DR WPI; 1999-181032/15.
 XX
 PT Scavenger receptor proteins - for treatment and diagnosis of
 PT disorders involving cell stress
 XX
 PS Example 3; Page 114-115; 175pp; Japanese.
 XX
 CC This sequence represents a fragment of a DNA encoding a human cellular
 CC stress response (CSR) protein of the invention. The CSR proteins are
 CC macrophage scavenger receptor proteins. The CSR proteins can be used in
 CC the treatment, gene therapy and diagnosis of diseases in which
 CC intracellular stress is important, such as arteriosclerosis, diabetic
 CC circulatory obstruction, and microbial infection. Expression of the
 CC proteins is induced in vivo in response to intracellular stress, and
 CC inhibits cell death as a result of such stress.
 XX
 SQ Sequence 525 BP; 111 A; 158 C; 151 G; 104 T; 1 other;

Query Match 2.3%; Score 17; DB 20; Length 525;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 gtcagttccatgttct 122
 Db 211 gtcagttccatgttct 227
 |||||

RESULT 5
 AAZ87633/C
 ID AAZ87633 standard; DNA; 824 BP.
 XX
 AC AAZ87633;
 XX
 DT 04-MAY-2000 (first entry)
 XX
 DE Sheep beta-lactoglobulin (betaLG) gene fragment.
 XX
 KW Transgenic bovine; transgene; milk; serum protein; industrial enzyme;
 KW infant formulation; lactoferrin; intestinal tract infection; lysozyme;
 KW iron absorption; albumin; antibacterial; iron sequestration; betaLG;
 KW beta-lactoglobulin; sheep; ds.
 XX
 OS Ovis sp.
 XX
 PN US6013857-A.
 XX
 PD 11-JAN-2000.
 XX
 PF 05-JUN-1995; 95US-0464167.
 XX
 PR 16-NOV-1993; 93US-0154019.
 PR 01-DEC-1989; 89US-0444745.
 PR 27-NOV-1990; 90US-0619131.
 PR 15-JUN-1992; 92US-0898956.
 PR 15-JUN-1993; 93US-0077788.
 XX
 PA (PHAR-) PHARMING BV.
 XX
 PI Deboer HA, Heyneker HL, Platenburg G, Krimpenfort PJA, Lee SH;
 PI Pieper F, Strijker R;
 XX
 DR WPI; 2000-146563/13.
 XX
 PT Transgenic cattle containing transgene controlled by mammary-specific
 PT regulator, for expressing proteins in the milk, particularly human
 PT lactoferrin for infant feeding formulations -
 XX
 PS Example 20; Fig 24; 92pp; English.
 XX
 CC The invention provides a transgenic bovine in which the somatic and germ
 CC cells contain a transgene comprising a regulatory sequence from a gene
 CC expressed in mammary glands, DNA encoding a signal sequence and DNA
 CC encoding a naturally occurring heterologous polypeptide. The transgenic
 CC bovine, or its descendants, produce milk containing the heterologous
 CC polypeptide. The transgenic bovines are used to express human milk and
 CC serum proteins or industrial enzymes, specifically for infant
 CC formulations that contain human lactoferrin for control of intestinal
 CC tract infections and to improve iron absorption, particularly when
 CC potentiated by human lysozyme. The polypeptide expressed may also be
 CC human albumin, used as a plasma extender. The polypeptide expressed in
 CC milk of the transgenic bovine requires little if any purification before
 CC human consumption and is expressed at significantly higher levels than in
 CC transgenic mice or sheep. Large polypeptides that are difficult to
 CC express in other systems can also be expressed. The present sequence
 CC represents a sheep beta-lactoglobulin gene fragment.
 XX
 SQ Sequence 824 BP; 138 A; 278 C; 245 G; 163 T; 0 other;

Query Match 2.3%; Score 17; DB 21; Length 824;
 Best Local Similarity 100.0%; Pred. No. 49;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcaagggtgac 606
 |||||
 Db 281 GCTGGGGCAAGGTGAC 265

RESULT 6
 AAC68328/c
 ID AAC68328 standard; DNA; 824 BP.

XX AC AAC68328;

XX DT 20-FEB-2001 (first entry)

XX DE Sheep beta lactoglobulin DNA.

XX KW Lactoferrin; mammary; milk; ss..

XX OS Ovis aries.

XX XX US6140552-A.

XX PN 31-OCT-2000.

XX PD 07-JUN-1995; 95US-0476798.

XX PR 16-NOV-1993; 93US-0154019.

XX PR 01-DEC-1989; 89US-0444745.

XX PR 27-NOV-1990; 90US-0619131.

XX PR 15-JUN-1992; 92US-0898956.

XX PR 15-JUN-1993; 93US-0077788.

XX PA (PHAR-) PHARMING BV.

XX PI Strijker R, Heyneker HL, Platenburg G, Pieper F, Krimpenfort PJA;

XX PI Lee SH, Deboer HA;

XX XX WPI; 2001-040323/05.

XX DR New transgenic bovine whose mammary gland cells contain DNA encoding a

XX PT signal sequence, and a polypeptide of interest and an expression

XX PT regulatory sequence, for producing polypeptides in bovine milk -

XX XX Disclosure; Fig 24; 88pp; English.

XX CC The present invention relates to a transgenic or chimeric bovine whose

XX CC mammary gland cells contain a construct encoding a signal sequence, a

XX CC polypeptide of interest and a regulatory sequence that promotes

XX CC expression of the DNA sequence. The transgenic or chimeric bovine is

XX CC useful for producing recombinant polypeptides in milk of female

XX CC transgenic mammals. The recombinant polypeptide may be used in food

XX CC formulations, particularly in infant formula having either

XX CC nutritional or beneficial value. An infant formula containing human

XX CC lactoferrin from the transgenic bovine milk provides bacteriostatic

XX CC effect, which aids in controlling diarrhoea in newborn. Recombinant

XX CC polypeptides may also be used to supplement common diet formulations.

XX XX Sequence 824 BP; 138 A; 278 C; 245 G; 163 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 824;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcaagggtgac 606

|||||

Db 281 GCTGGGGCAAGGTGAC 265

RESULT 7

AAV69855/c

ID AAV69855 standard; DNA; 3564 BP.

XX AC

XX AAV69855;

XX DT 28-JAN-1999 (first entry)

XX DE Human Y chromosome fragment #4.

XX KW Human; Y chromosome; in situ hybridisation; probe; gender; prenatal;

XX KW forensic investigation; ss.

XX OS Homo sapiens.

XX PN US5840482-A.

XX PD 24-NOV-1998.

XX PF 10-OCT-1990; 90US-0594921.

XX PR 10-OCT-1990; 90US-0594921.

XX XX (REGC) UNIV CALIFORNIA.

XX PI Gray JW, Weier H;

XX DR WPI; 1999-034027/03.

XX PT Probes for detecting Y chromosome DNA sequence - useful for gender,

XX PT prenatal or forensic investigations

XX PS Example 1; Fig 1; 28pp; English.

XX CC The present invention describes probes for detecting a highly repeated

XX CC human Y chromosome DNA sequence. The present sequence represents a

XX CC fragment of the Y chromosome DNA to parts of which a probe could

XX CC hybridise. The probes are useful for: (1) gender determination where a

XX CC family history indicates the possibility of a sex-linked disease e.g.

XX CC haemophilia, Lesch-Nyhan syndrome, Fabry disease or Duchenne muscular

XX CC dystrophy; (2) prenatal diagnosis e.g. to enable obstetricians to

XX CC instigate treatment regimes based on gender as female foetuses have

XX CC greater survivability compared to male foetuses of the same size and

XX CC gestational age, and also to permit first trimester abortion in families

XX CC with a strong gender preference particularly in countries where

XX CC opportunities for women are very limited and where the infanticide of

XX CC female newborns is commonly practised; and (3) in forensic investigations

XX CC to detect the presence of severely decayed sperm, and to type for gender

XX CC hair and fragmentary tissue samples.

XX SQ Sequence 3564 BP; 742 A; 1081 C; 298 G; 1443 T; 0 other;

Query Match 2.3%; Score 17; DB 20; Length 3564;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 300 ctacggtctcgaatgga 316

|||||

Db 3507 CTACGGTCTCGAATGGA 3491

RESULT 8

AAV69855/c

ID AAV69855 standard; DNA; 3760 BP.

XX AC AAV69855;

XX DT 04-MAY-1999 (first entry)

XX DE Borrelia burgdorferi polynucleotide sequence #32.

XX KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

XX KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

XX KW infection; diagnosis; characterisation; detection; ds.

```

OS Borrelia burgdorferi.
PN WO9858943-A1.
XX
XX
XX 30-DEC-1998.
XX
XX 18-JUN-1998; 98WO-US12764.
XX
XX 03-SEP-1997; 97US-0057483.
XX
XX 20-JUN-1997; 97US-0050359.
XX
XX 22-JUL-1997; 97US-0053344.
XX
XX 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX (MEDI-) MEDIMMUNE INC.
XX
XX Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
XX White OR;
XX
XX WPI; 1999-081217/07.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the detection, diagnosis, characterisation, prevention
XX and therapy of infections, particularly Lyme disease
XX
XX Claim 1; Page 991-993; 1128pp; English.
XX
XX AAX20248 to AAX20402 represent polynucleotide sequences isolated from
XX Borrelia burgdorferi (Bb). Products derived from Bb can be used for
XX the detection, diagnosis, characterisation, prevention and therapy of
XX Bb infections, e.g. Lyme disease. They can also be used for the
XX production of biosynthetic products, e.g. enzymes. Borrelia belongs
XX to a family of motile, spiral-shaped bacteria called Spirochetes.
XX Spirochetes are pathogenic in humans and Borrelia causes epidemic and
XX endemic relapsing fever, and Lyme borreliosis, more commonly known as
XX Lyme disease.
XX
XX Sequence 3760 BP; 1477 A; 435 C; 435 G; 1412 T; 1 other;

Query Match 2.3%; Score 17; DB 20; Length 3760;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 aagggttataagtataac 515
   |||||
Db 3075 AAGCTTTATAAGTTATAC 3059

RESULT 9
AAT79728/c
ID AAT79728 standard; DNA; 6763 BP.
XX
XX AAT79728;
XX
XX 11-DEC-1997 (first entry)
XX
XX Beta-lactoglobulin cloning vector pMAD6.
XX
XX Protein C; transgenic animal; sheep; rabbit; cattle; goat; milk;
XX blood clotting; anticoagulant; beta-lactoglobulin; pMAD6; vector;
XX ss.
XX
XX Chimeric Ovis aries;
XX
XX Chimeric synthetic.
XX
XX WO9720043-A1.
XX
XX 05-JUN-1997.
XX
XX 26-NOV-1996; 96WO-US18866.
XX
XX 13-JUN-1996; 96US-0019692.

```

```

PR 30-NOV-1995; 95US-0565074.
XX
XX (PPLT-) PPL THERAPEUTICS.
XX
XX (ZYMO ) ZYMOGENETICS INC.
XX
XX Cottingham I, Foster DC, Garner I, Prunkard DE;
XX Sprecher CA, Temperley SM;
XX
XX WPI; 1997-310599/28.
XX
XX Production of protein C in transgenic animal - useful for high
XX quantity protein C production with therapeutic value
XX
XX Example 1; Page 80-85; 99pp; English.
XX
XX This DNA sequence comprises plasmid pMAD6, a pUC18-based vector
XX containing ovine beta-lactoglobulin gene sequences (see AAT79725).
XX Nucleic acids encoding two-chain cleavage site-modified human
XX protein C (see AAT79723-24) can be cloned into the vector so that
XX the protein C gene is under the control of the beta-lactoglobulin
XX promoter. This construct can be used in claimed methods for the
XX production of human protein C in the milk of transgenic animals.
XX Modification of the protein C two-chain cleavage site improves the
XX maturation of single-chain protein C to two-chain protein C.
XX
XX Sequence 6763 BP; 1473 A; 1817 C; 1709 G; 1764 T; 0 other;

Query Match 2.3%; Score 17; DB 18; Length 6763;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcaagggtgac 606
   |||||
Db 3774 GCTGGGGCAAGGGTGAC 3758

RESULT 10
AAV74380/c
ID AAV74380 standard; DNA; 7494 BP.
XX
XX AAV74380;
XX
XX 16-MAR-1999 (first entry)
XX
XX Staphylococcus aureus contig SEQ ID #69.
XX
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
XX
XX Staphylococcus aureus.
XX
XX Key Location/Qualifiers
XX misc_feature 901..960
XX /*tag= a
XX /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX misc_feature 2701..2760
XX /*tag= b
XX /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX misc_feature 4501..4560
XX /*tag= c
XX /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX

```

```
FT misc_feature 6301..6360
FT /*tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX
XX
XX
XX EP786519-A2.
XX
XX
XX PD 30-JUL-1997.
XX
XX
XX PF 07-JAN-1997; 97EP-0100117.
XX
XX PR 05-JAN-1996; 96US-0009861.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX Barash SC, Choi GH, Dillon FU, Fannon MR, Kunsch CA;
XX Rosen CA;
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
XX stored on computer readable medium and used in the production of
XX anti-S.aureus vaccines
XX
XX Claim 1; Page 503-507; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX of the invention. The DNA sequences are recorded on a computer readable
XX medium, preferably selected from a floppy or hard disk, random access
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX the S.aureus DNA sequences allows putative functions to be assigned so
XX that protein-encoding or regulatory regions of commercial, therapeutic or
XX industrial importance can be obtained. Specifically, sequences which are
XX likely to encode antigens have been identified and these polypeptides can
XX be used in a vaccine composition against S.aureus infection. The
XX polypeptides can also be used in a kit for the immunodetection of
XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX skin and surgical wound infections, scalded skin syndrome, toxic shock
XX syndrome, etc. Organisms transformed with the DNA sequences can be used
XX for recombinant production of the polypeptides. The new DNA sequences
XX (and their fragments) are useful as primers or probes for isolating
XX homologues of any of the S.aureus DNA sequences contained on the
XX computer readable medium.
XX
XX Sequence 7494 BP; 2436 A; 1309 C; 1029 G; 2473 T; 247 other;

Query Match 2.3%; Score 17; DB 18; Length 7494;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 aagtgatattgaagtt 183
| | | | | | | | | | | | | | | | | | | | | |
DB 1535 AAGTGGATATTGCACTT 1519

RESULT 11
AAT03855/c
ID AAT03855 standard; DNA; 10807 BP.
XX
XX AC AAT03855;
XX
XX DT 26-OCT-1996 (first entry)
XX
XX DE Sheep beta-lactoglobulin coding sequence.
XX
XX KW Sheep beta-lactoglobulin; signal peptide; promoter; fusion protein;
XX human fibrinogen A-alpha chain; B-beta chain; gamma chain;
XX transgenic animal milk; ds.

Query Match 2.3%; Score 17; DB 16; Length 10807;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcgaagggtgac 606
| | | | | | | | | | | | | | | | | | | | | |
DB 3776 GCTGGGCAAGGGTGAC 3760

RESULT 12
AAT79725/c
ID AAT79725 standard; DNA; 10807 BP.
XX
XX AC AAT79725;
XX
XX DT 11-DEC-1997 (first entry)
XX
XX DE Ovine beta-lactoglobulin gene.
XX
XX KW Protein C; transgenic animal; sheep; rabbit; cattle; goat; milk;
XX blood clotting; anticoagulant; beta-lactoglobulin; promoter; ss.
XX
XX OS Ovis aries.
XX
XX FH Key 1.4257
XX 5'UTR /*tag= a
XX
XX FT W05720043-A1.
XX
XX PD 05-JUN-1997.
```

```
OS Ovis aries.
XX
XX PN W09523868-A1.
XX
XX PD 08-SEP-1995.
XX
XX PF 01-MAR-1995; 95WO-US02648.
XX
XX PR 03-MAR-1994; 94US-0206176.
XX
XX PA (PHAR-) PHARM PROTEINS LTD.
XX (ZYMO ) ZYMOGENETICS INC.
XX
XX PI Dairymple MA, Foster DC, Garner I, Prunkard DE;
XX WPI; 1995-320582/41.
XX
XX Production of fibrinogen in transgenic mammals - by introducing DNA
XX segments into the germ line of a non-human mammal and collecting
XX milk from female progeny.
XX
XX Disclosure; Page 69-77; 99pp; English.
XX
XX This sequence encodes sheep beta-lactoglobulin and contains a unique
XX EcoRV site at nucleotide 4245 in the 5' untranslated region of the
XX gene. This site allows insertion of additional DNA sequences under
XX the control of the beta-lactoglobulin promoter 3' to the
XX transcription initiation site. The promoter and signal peptide may
XX be used for mamma tissue-specific gene expression and secretion of
XX human fibrinogen A-chain (see AAT03853), B-beta chain (AAT03852) and
XX gamma chain (AAT03854) in non-human transgenic animals. A region of
XX at least the proximal 406 bp of 5' flanking sequences (nucleotides
XX 3844-4257) is generally used, although larger sequences such as
XX nucleotides 1-4257 are preferred. Furthermore, the region
XX surrounding the initiation ATG of one or more of the human
XX fibrinogen sequences may be replaced with the corresponding
XX beta-lactoglobulin sequences providing a putative tissue-specific
XX initiation environment to enhance expression.
XX
XX Sequence 10807 BP; 2215 A; 3065 C; 2985 G; 2542 T; 0 other;

Query Match 2.3%; Score 17; DB 16; Length 10807;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcgaagggtgac 606
| | | | | | | | | | | | | | | | | | | | | |
DB 3776 GCTGGGCAAGGGTGAC 3760

RESULT 12
AAT79725/c
ID AAT79725 standard; DNA; 10807 BP.
XX
XX AC AAT79725;
XX
XX DT 11-DEC-1997 (first entry)
XX
XX DE Ovine beta-lactoglobulin gene.
XX
XX KW Protein C; transgenic animal; sheep; rabbit; cattle; goat; milk;
XX blood clotting; anticoagulant; beta-lactoglobulin; promoter; ss.
XX
XX OS Ovis aries.
XX
XX FH Key 1.4257
XX 5'UTR /*tag= a
XX
XX FT W05720043-A1.
XX
XX PD 05-JUN-1997.
```

```

XX PF 26-NOV-1996; 96WO-US18866.
XX XX
XX PR 13-JUN-1996; 96US-0019692.
XX PR 30-NOV-1995; 95US-0565074.
XX XX
XX PA (PPLT-) PPL THERAPEUTICS.
XX PA (Zymo ) ZYMOGENETICS INC.
XX XX
XX PI Cottingham I, Foster DC, Garner I, Prunkard DE;
XX PI Sprecher CA, Temperley SM;
XX XX
XX DR WPI; 1997-310599/28.
XX PT
XX PT Production of protein C in transgenic animal - useful for high
XX PT quantity protein C production with therapeutic value
XX XX
XX PS Disclosure; Page 66-74; 99pp; English.
XX XX
XX CC This DNA sequence comprises the ovine beta-lactoglobulin gene
XX CC including the promoter region. A DNA segment encompassing the
XX CC 5' flanking promoter region can be utilised in novel constructs
XX CC for the expression of human protein C in the milk of a transgenic
XX CC animal. A claimed method involves: (a) providing a DNA construct
XX CC comprising DNA encoding a secretion signal and a protein C
XX CC propeptide, operably linked to DNA encoding two-chain cleavage
XX CC site-modified protein C (see AAW25085-86), the 2 DNA sequences being
XX CC linked to elements required for protein C expression in a mammary
XX CC gland of a host female animal, such as beta-lactoglobulin gene
XX CC promoter; and (b) using the DNA construct to breed a transgenic
XX CC animal (esp. sheep, rabbit, cattle, goat) that produces protein C
XX CC in its milk, at least 90% of the protein C being in the two-chain
XX CC form.
XX XX
XX SQ Sequence 10807 BP; 2214 A; 3066 C; 2985 G; 2542 T; 0 other;

Query Match 2.3%; Score 17; DB 18; Length 10807;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctgggggcaagggtgac 606
DB 3776 GCTGGGGCAAGGGTGAC 3760
|||||
RESULT 13
AAF28652/c
ID AAF28652 standard; DNA; 12061 BP.
XX AC
XX AC AAF28652;
XX XX
XX DT 03-APR-2001 (first entry)
XX XX
XX DE Partial DNA sequence #1 of pCLYSM.
XX XX
XX KW Transgenic; milk; lysozyme; pCLYSM; human; salmon; calcitonin; ovine;
XX KW beta-lactoglobulin promoter; ss.
XX XX
XX OS Unidentified.
XX XX
XX PN WO200100855-A1.
XX XX
XX PD 04-JAN-2001.
XX XX
XX PF 23-JUN-2000; 2000WO-GB02459.
XX XX
XX PR 23-JUN-1999; 99GB-0014733.
XX PR 10-AUG-1999; 99US-0147819.
XX XX
XX PA (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.
XX XX
XX PI Cottingham IR, McCreath GE;

```

```

XX WPI; 2001-137958/14.
XX XX
XX PT Producing peptides for use as medicaments, nutritional additives or
XX PT research tools, comprises expressing a fusion protein, which contains
XX PT the peptide linked to a lysozyme fusion partner protein, in the milk of
XX PT a transgenic mammal.
XX XX
XX PS Example 1; Fig 2; 57pp; English.
XX XX
XX CC The present invention relates to a method for producing a peptide. The
XX CC method comprises expressing in the milk of a transgenic non-human
XX CC placental mammal a fusion protein, comprising the peptide to be expressed
XX CC linked to lysozyme (a fusion partner protein). The method is useful for
XX CC producing proteins or peptides that are useful as medicaments,
XX CC nutritional additives or research tools. The present sequence is a
XX CC partial DNA sequence of pCLYSM. pCLYSM was designed to express a human
XX CC lysozyme-salmon calcitonin fusion protein in the milk of transgenic
XX CC animals. The fusion protein allows the release of calcitonin from the end
XX CC of a linker arm fused to the lysozyme C terminal by cyanogen bromide
XX CC chemical cleavage. pCLYSM consists of the ovine beta-lactoglobulin (BLG)
XX CC promoter and 5' UTR, human lysozyme gene, peptide linker (see AAB61572),
XX CC CNBR cleavage site and salmon calcitonin coding sequence, ovine BLG 3',
XX CC UTR, polyadenylation site and flanking region, chick alpha-globin
XX CC insulator region and the pUC18 bacterial plasmid vector. The present
XX CC sequence is the DNA sequence of pCLYSM as described above, but excluding
XX CC the bacterial plasmid sequence.
XX XX
XX SQ Sequence 12061 BP; 3029 A; 2803 C; 2777 G; 3452 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 12061;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctgggggcaagggtgac 606
DB 3792 GCTGGGGCAAGGGTGAC 3776
|||||
RESULT 14
AAT47715
ID AAT47715 standard; DNA; 15144 BP.
XX AC
XX AC AAT47715;
XX XX
XX DT 20-MAY-1997 (first entry)
XX XX
XX DE Mouse bone morphogenetic protein-2 gene and promoter region.
XX XX
XX KW Osteogenic agent; bone morphogenetic protein-2; BMP-2;
XX KW growth factor; osteoblast; promoter; osteoporosis; fracture repair;
XX KW osteoblastic metastasis; osteosclerosis; therapy; ss.
XX XX
XX OS Mus sp.
XX XX
XX PN WO9638590-A1.
XX XX
XX PD 05-DEC-1996.
XX XX
XX PF 31-MAY-1996; 96WO-US08197.
XX XX
XX PR 02-JUN-1995; 95US-0458434.
XX XX
XX PA (OSTE-) OSTEOSCREEN INC.
XX XX
XX PI Feng JQ, Ghosh-Choudhury N, Harris SE, Mundy GR;
XX XX
XX XX WPI; 1997-034396/03.
XX XX
XX PT System for identifying osteogenic agents that induce prodn. of bone
XX PT morphogenetic protein - is cell contg. reporter gene under control of
XX PT BMP gene promoter, also new promoters of BMP-2 and -4 and related

```

PT vectors and cells
 PS Example 2; Page 33-42; 76pp; English.
 XX
 CC A DNA sequence (AA747715) comprises the mouse morphogenetic protein-2
 CC (BMP-2) gene and promoter regions. It was isolated from a mouse
 CC genomic library (B6/CBA) using mouse BMP-2 cDNA as probe. Analysis
 CC of the 5' flanking sequence (see also AA747713) of the gene revealed
 CC putative response elements and transcription factor recognition
 CC sequences. The promoter region of the BMP-2 or BMP-4 (see also
 CC AA747711) gene can be linked to a reporter gene in a recombinant
 CC expression vector. Host cells transfected with the vector can be
 CC used to identify novel osteogenic agents that stimulate bone cells
 CC to produce BMP family bone growth factors. Such osteogenic agents
 CC can be used to promote bone growth and healing.
 XX
 SQ Sequence 15144 BP; 4062 A; 3380 C; 3503 G; 4199 T; 0 other;
 Query Match 2.3%; Score 17; DB 18; Length 15144;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 tggaaggccctgggtag 166
 Db 8938 tggaaggccctgggtag 8954
 |||||

RESULT 15
 AAA53077
 ID AAA53077 standard; DNA; 33 BP.
 XX
 AC AAA53077;
 XX
 DT 15-SEP-2000 (first entry)
 XX
 DE Rat genomic DNA vector pSP72 oligonucleotide Tag 1 SEQ ID NO:1.
 XX
 KW Genome analysis; tag; chromosomal location; integration site;
 KW insertion element; identification; polymorphism; PCR; primer; ss.
 XX
 OS Rattus sp.
 OS Synthetic.
 XX
 PN WO200024937-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 27-OCT-1999; 99WO-US25037.
 XX
 PR 28-OCT-1998; 98US-0105914.
 PR 26-OCT-1999; 99US-0427834.
 XX
 PA (STRA/) STRATHMANN M.
 XX
 PI Strathmann M;
 XX
 DR WPI; 2000-350769/30.
 XX

Search completed: August 23, 2001, 15:01:15
 Job time: 9833 sec

Query Match 2.2%; Score 16; DB 21; Length 33;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 ggaagggtgcccaggtt 381
 Db 9 ggaagggtgcccaggtt 24
 |||||

CC comparing them to a sequence database; (5) identifying a gene associated
 CC with a phenotype by assaying a distribution of polymorphisms in a
 CC population to identify a subset of the polymorphisms that associate with
 CC the phenotype and identifying the gene linked to the subset of
 CC polymorphisms or by locating the gene to within a defined genomic region
 CC and sequencing the region; (6) identifying a compound that interacts
 CC with a gene product or polypeptide or modulates expression of a gene;
 CC (7) preparing antibodies to a polypeptide; (8) preparing an array of
 CC oligonucleotides for assaying polynucleotides of known sequence; (9)
 CC producing a transgenic cell; (10) producing a cell with a mutated gene
 CC by sequencing and identifying a gene, introducing a mutation into the
 CC gene in the cell; (11) assaying gene expression; (12) identifying a
 CC bioactive polypeptide; (13) locating genome rearrangements by generating
 CC physical maps of polynucleotides and comparing them to a database of
 CC physical maps and identifying differences in landmarks to locate the
 CC genomic rearrangements; and (14) identifying a gene associated with a
 CC disease. Cells produced can be used to generate transgenic organisms
 CC which are then used to identify bioactive compounds. AAA53077 to AAA53125
 CC represent oligonucleotide sequences used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 33 BP; 8 A; 7 C; 13 G; 5 T; 0 other;

Query Match 2.2%; Score 16; DB 21; Length 33;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 ggaagggtgcccaggtt 381
 Db 9 ggaagggtgcccaggtt 24
 |||||

PT Parallel methods of genomic analysis useful for determining
 PT polymorphisms, chromosome rearrangements and generating physical maps -
 PS Example 1; Page 91; 153pp; English.
 XX

CC The present invention describes parallel methods of genomic analysis
 CC using polynucleotides associated with sample tags. The parallel methods
 CC are used for the following claimed procedures: (1) constructing a
 CC recombinant molecule consisting of a sequence element from a homologue
 CC of a sequenced polynucleotide joined to a vector; (2) producing a
 CC polypeptide, preferably of known sequence, from a recombinant molecule
 CC as in (1) which has been transferred to a host; (3) constructing a
 CC database of genetic information from polynucleotides sequenced using the
 CC method; (4) identifying polymorphisms by sequencing polynucleotides and

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 11:45:08 ; Search time 1867.27 Seconds
(without alignments)
3766.420 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgtctcg 744

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_estl1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estomi:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: em_est100:*
170: em_est101:*
171: em_est102:*
172: em_est103:*
173: em_est104:*
174: em_est105:*
175: em_est106:*
176: em_est107:*
177: em_est108:*
178: em_est109:*
179: em_est110:*
180: em_est111:*
181: em_est112:*
182: em_est113:*
183: em_est114:*
184: em_est115:*
185: em_est116:*
186: em_est117:*
187: em_est118:*
188: em_est119:*
189: em_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

Prd. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20	2.7	839	146	BF261375	HV_CEA000
C 2	19	2.6	517	226	AQ285087	RPC111-79
C 3	19	2.6	559	242	AZ361905	IM0106J15
C 4	19	2.6	626	258	L6474AX	AL354335 Leishmani
C 5	19	2.6	636	228	AQ427471	CITBI-E1-
C 6	18	2.4	275	29	AV357168	AV357168 AV357168
C 7	18	2.4	299	172	BG012371	RC2-GN028
C 8	18	2.4	332	243	AZ426939	IM0208H08
C 9	18	2.4	342	107	AU055708	AU055708 AU055708
C 10	18	2.4	362	15	AI069694	TENU2738
C 11	18	2.4	371	21	AI544667	fb77d09.x
C 12	18	2.4	429	115	AW390681	CM2-ST019
C 13	18	2.4	436	235	AQ88579	HS_5544_A
C 14	18	2.4	446	137	BE06626	WHE0901_G
C 15	18	2.4	448	148	BF422269	FMI_13_D0
C 16	18	2.4	454	17	AI186540	qd35d03.x
C 17	18	2.4	461	167	BE404293	WHE1204_E
C 18	18	2.4	464	234	AQ862731	nbeb0019M
C 19	18	2.4	469	232	AQ724042	HS_5375_A
C 20	18	2.4	470	8	AA542472	fa07e08.s
C 21	18	2.4	488	5	AA308387	EST179219
C 22	18	2.4	502	167	BE422602	WHE0055_A
C 23	18	2.4	509	145	BF202269	WHE0984_F
C 24	18	2.4	511	166	BE355369	DGL_114_C
C 25	18	2.4	523	166	BE355291	DGL_114_C
C 26	18	2.4	527	146	BF259611	HVSMET001
C 27	18	2.4	551	258	TA7G12Q	AL452262 T. brucei
C 28	18	2.4	562	258	TA19G07P	AL451891 T. brucei
C 29	18	2.4	580	243	AZ426948	LM0208J08
C 30	18	2.4	586	24	AI729124	BNLGH126
C 31	18	2.4	592	122	AW961291	EST373363
C 32	18	2.4	592	234	AQ865167	nbeb0024J
C 33	18	2.4	605	105	AL036448	DKFZp564D
C 34	18	2.4	606	146	BF291791	WHE2204_A
C 35	18	2.4	607	167	BE423985	WHE0066_F
C 36	18	2.4	620	231	AQ657042	Sheared D
C 37	18	2.4	626	245	AZ519171	RPC1-11-7
C 38	18	2.4	638	233	AQ754786	HS_3166_B
C 39	18	2.4	644	166	BE356747	DGL_12_B0
C 40	18	2.4	696	16	AI110205	GH09761.5
C 41	18	2.4	702	223	AQ013977	RPC111-25
C 42	18	2.4	753	141	BE885533	601508727
C 43	18	2.4	822	153	BG418473	HVSMET002
C 44	18	2.4	825	233	AQ4747536	HS_5536_A
C 45	18	2.4	867	146	BF267575	HV_CEA001

ALIGNMENTS

RESULT 1	BF261375	839 bp	EST	09-MAR-2001
LOCUS	HV_CEA0001B22f	Hordeum vulgare seedling green leaf EST library		
DEFINITION	HVCDA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone			
ACCESSION	HV_CEA0001B22f	mRNA sequence.		
VERSION	BF261375			
KEYWORDS	BF261375.2	GI:13259773		
SOURCE	barley.			
ORGANISM	Hordeum vulgare			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae			
	; Triticeae; Hordeum.			
	1 (bases 1 to 839)			
	Wing, R., Close, T.J., Klein, H., Wise, R., Begum, D., Frisch, D., Yu			
	, Y., Anderson, H., Dale, J., Henry, D., Kern, D., Palmer, M., Rambo			

TITLE	T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.				
JOURNAL	Development of a genetically and physically anchored EST resource for barley genomics				
COMMENT	Unpublished (2000)				
	On Nov 17, 2000 this sequence version replaced gi:11192362.				
	Contact: Wing RA				
	Clemson University Genomics Institute				
	Clemson University				
	100 Jordan Hall, Clemson, SC 29634, USA				
	Tel: 864 656 7288				
	Fax: 864 656 4293				
	Email: rwing@clemson.edu				
	Seq primer: AATTAACCTCTACTAAAGGG				
	High quality sequence stop: 719.				
FEATURES	Location/Qualifiers				
	1..839	/organism="Hordeum vulgare"			
		/cultivar="C116155 (Mial3)"			
		/db_xref="taxon:4513"			
		/clone_lib="Hordeum vulgare seedling green leaf EST"			
		/library_HVCDA0004 (Erysiphe infected & control)"			
		/tissue_type="seedling green leaf"			
		/lab_host="TJCI21"			
		/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"			
BASE COUNT	159 a	247 c	225 g	208 t	
ORIGIN					
	Query Match 2.7%; Score 20; DB 146; Length 839;				
	Best Local Similarity 100.0%; Pred. No. 5.3;				
	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	452	agtcgatgaatcaagcctt	471		
Db	739	AGTCGATGAATCAAGCCTT	720		
RESULT 2	AQ285087	517 bp	DNA	GSS	04-MAY-1999
LOCUS	RPC111-79N17.TJ	RPC1-11	Homo sapiens	genomic clone	RPC1-11-79N17,
DEFINITION	DNA sequence.				
ACCESSION	AQ285087				
VERSION	AQ285087.1	GI:3911327			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 517)				
AUTHORS	Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.				
TITLE	Use of human BAC End Sequences for Sequence-Ready Map Building				
JOURNAL	Unpublished (1998)				
COMMENT	Other GSSs: RPC111-79N17.TJ				
	Contact: Mark Adams				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: mdadams@tigr.org				
	Clones are derived from the human BAC library RPC1-11. For BAC				
	library availability, please contact Pieter de Jong				
	(pieter@dejong.med.buffalo.edu). Clones may be purchased from				
	BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from				
	Research Genetics (info@resgen.com). BAC end search page:				
	http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html				
	Seq primer: SP6				
	Class: BAC ends.				
FEATURES	Location/Qualifiers				

source

1. .517
 /organism="Homo sapiens"
 /db_xref="GDB:7530280"
 /db_xref="taxon:9606"
 /clone="RPC1-11-79N17"
 /clone_lib="RPC1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPC111 Human Male BAC Library" 1 others
 BASE COUNT 168 a 97 c 54 g 197 t
 ORIGIN

Query Match 2.6%; Score 19; DB 226; Length 517;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 tcagttccatgttcctcta 125
 |||||
 Db 470 TCAGTTCATGTTCCCTCA 488

RESULT 3

AZ361905/c
 LOCUS 559 bp DNA 02-OCT-2000
 DEFINITION 1M0106J15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0106J15 R, DNA sequence.
 ACCESSION AZ361905
 VERSION AZ361905.1 GI:10475605
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 559)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0106 row: J column: 15
 Seq primer: CACACGAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 559.

FEATURES

source

1. 559
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0106J15"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (gi14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
 177 a 103 c 114 g 165 t

Query Match 2.6%; Score 19; DB 242; Length 559;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 aacttgacaggtacacagg 403
 |||||
 Db 423 AACTTGACAGGTACACAGG 405

RESULT 4

L6474AX 626 bp DNA GSS 29-APR-2000
 LOCUS Leishmania major Friedlin cosmid L6474.1 t3hyg end-sequence,
 genomic survey sequence.
 DEFINITION
 ACCESSION AL354335
 VERSION AL354335.1 GI:7671978
 KEYWORDS GSS.
 SOURCE Leishmania major.
 ORGANISM Leishmania major
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 626)
 Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
 Smith,D.F.
 TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 REFERENCE 2 (bases 1 to 626)
 TAYLOR,R.G., HUCKLE,E.E.J., IVENS,A.C., RAJANDREAM,M.A. and
 BARRELL,B.G.
 TITLE Direct Submission
 JOURNAL Submitted (28-APR-2000) Leishmania major Friedlin genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 alicat@sanger.ac.uk
 COMMENT see http://www.ebi.ac.uk/parasites/leish.html
 Details of Leishmania sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/Projects/L_major/
 The cLHYG t3hyg primer sequence can be obtained from acc. no.
 U59231.

FEATURES

source

1. 626
 Location/Qualifiers
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="cosmid L6474.1"

BASE COUNT 96 a 213 c 186 g 129 t 2 others
 ORIGIN

Query Match 2.6%; Score 19; DB 258; Length 626;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 cagggcgaagcgccagcg 541
 |||||
 Db 79 CAGGGCGAAGCGCGCAGCG 97

```

RESULT 5
AQ427471/c
LOCUS      636 bp      DNA      GSS      24-MAR-1999
DEFINITION CITBI-E1-2574N14.TF CITBI-E1 Homo sapiens genomic clone 2574N14,
DNA sequence.
ACCESSION  AQ427471
VERSION     AQ427471.1 GI:4500376
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 636)
AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
JOURNAL   Map Building
COMMENT   Unpublished (1997)
Other_GSSs: CITBI-E1-2574N14.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..636
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2574N14"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
BASE COUNT          278 a 126 c 117 g 115 t
ORIGIN
111 ttccatgttcctctaccag 129
112 |
113 |
114 |
115 |
116 |
117 |
118 |
119 |
120 |
121 |
122 |
123 |
124 |
125 |
126 |
127 |
128 |
129 |
130 |
131 |
132 |
133 |
134 |
135 |
136 |
137 |
138 |
139 |
140 |
141 |
142 |
143 |
144 |
145 |
146 |
147 |
148 |
149 |
150 |
151 |
152 |
153 |
154 |
155 |
156 |
157 |
158 |
159 |
160 |
161 |
162 |
163 |
164 |
165 |
166 |
167 |
168 |
169 |
170 |
171 |
172 |
173 |
174 |
175 |
176 |
177 |
178 |
179 |
180 |
181 |
182 |
183 |
184 |
185 |
186 |
187 |
188 |
189 |
190 |
191 |
192 |
193 |
194 |
195 |
196 |
197 |
198 |
199 |
200 |
201 |
202 |
203 |
204 |
205 |
206 |
207 |
208 |
209 |
210 |
211 |
212 |
213 |
214 |
215 |
216 |
217 |
218 |
219 |
220 |
221 |
222 |
223 |
224 |
225 |
226 |
227 |
228 |
229 |
230 |
231 |
232 |
233 |
234 |
235 |
236 |
237 |
238 |
239 |
240 |
241 |
242 |
243 |
244 |
245 |
246 |
247 |
248 |
249 |
250 |
251 |
252 |
253 |
254 |
255 |
256 |
257 |
258 |
259 |
260 |
261 |
262 |
263 |
264 |
265 |
266 |
267 |
268 |
269 |
270 |
271 |
272 |
273 |
274 |
275 |
276 |
277 |
278 |
279 |
280 |
281 |
282 |
283 |
284 |
285 |
286 |
287 |
288 |
289 |
290 |
291 |
292 |
293 |
294 |
295 |
296 |
297 |
298 |
299 |
300 |
301 |
302 |
303 |
304 |
305 |
306 |
307 |
308 |
309 |
310 |
311 |
312 |
313 |
314 |
315 |
316 |
317 |
318 |
319 |
320 |
321 |
322 |
323 |
324 |
325 |
326 |
327 |
328 |
329 |
330 |
331 |
332 |
333 |
334 |
335 |
336 |
337 |
338 |
339 |
340 |
341 |
342 |
343 |
344 |
345 |
346 |
347 |
348 |
349 |
350 |
351 |
352 |
353 |
354 |
355 |
356 |
357 |
358 |
359 |
360 |
361 |
362 |
363 |
364 |
365 |
366 |
367 |
368 |
369 |
370 |
371 |
372 |
373 |
374 |
375 |
376 |
377 |
378 |
379 |
380 |
381 |
382 |
383 |
384 |
385 |
386 |
387 |
388 |
389 |
390 |
391 |
392 |
393 |
394 |
395 |
396 |
397 |
398 |
399 |
400 |
401 |
402 |
403 |
404 |
405 |
406 |
407 |
408 |
409 |
410 |
411 |
412 |
413 |
414 |
415 |
416 |
417 |
418 |
419 |
420 |
421 |
422 |
423 |
424 |
425 |
426 |
427 |
428 |
429 |
430 |
431 |
432 |
433 |
434 |
435 |
436 |
437 |
438 |
439 |
440 |
441 |
442 |
443 |
444 |
445 |
446 |
447 |
448 |
449 |
450 |
451 |
452 |
453 |
454 |
455 |
456 |
457 |
458 |
459 |
460 |
461 |
462 |
463 |
464 |
465 |
466 |
467 |
468 |
469 |
470 |
471 |
472 |
473 |
474 |
475 |
476 |
477 |
478 |
479 |
480 |
481 |
482 |
483 |
484 |
485 |
486 |
487 |
488 |
489 |
490 |
491 |
492 |
493 |
494 |
495 |
496 |
497 |
498 |
499 |
500 |
501 |
502 |
503 |
504 |
505 |
506 |
507 |
508 |
509 |
510 |
511 |
512 |
513 |
514 |
515 |
516 |
517 |
518 |
519 |
520 |
521 |
522 |
523 |
524 |
525 |
526 |
527 |
528 |
529 |
530 |
531 |
532 |
533 |
534 |
535 |
536 |
537 |
538 |
539 |
540 |
541 |
542 |
543 |
544 |
545 |
546 |
547 |
548 |
549 |
550 |
551 |
552 |
553 |
554 |
555 |
556 |
557 |
558 |
559 |
560 |
561 |
562 |
563 |
564 |
565 |
566 |
567 |
568 |
569 |
570 |
571 |
572 |
573 |
574 |
575 |
576 |
577 |
578 |
579 |
580 |
581 |
582 |
583 |
584 |
585 |
586 |
587 |
588 |
589 |
590 |
591 |
592 |
593 |
594 |
595 |
596 |
597 |
598 |
599 |
600 |
601 |
602 |
603 |
604 |
605 |
606 |
607 |
608 |
609 |
610 |
611 |
612 |
613 |
614 |
615 |
616 |
617 |
618 |
619 |
620 |
621 |
622 |
623 |
624 |
625 |
626 |
627 |
628 |
629 |
630 |
631 |
632 |
633 |
634 |
635 |
636 |
637 |
638 |
639 |
640 |
641 |
642 |
643 |
644 |
645 |
646 |
647 |
648 |
649 |
650 |
651 |
652 |
653 |
654 |
655 |
656 |
657 |
658 |
659 |
660 |
661 |
662 |
663 |
664 |
665 |
666 |
667 |
668 |
669 |
670 |
671 |
672 |
673 |
674 |
675 |
676 |
677 |
678 |
679 |
680 |
681 |
682 |
683 |
684 |
685 |
686 |
687 |
688 |
689 |
690 |
691 |
692 |
693 |
694 |
695 |
696 |
697 |
698 |
699 |
700 |
701 |
702 |
703 |
704 |
705 |
706 |
707 |
708 |
709 |
710 |
711 |
712 |
713 |
714 |
715 |
716 |
717 |
718 |
719 |
720 |
721 |
722 |
723 |
724 |
725 |
726 |
727 |
728 |
729 |
730 |
731 |
732 |
733 |
734 |
735 |
736 |
737 |
738 |
739 |
740 |
741 |
742 |
743 |
744 |
745 |
746 |
747 |
748 |
749 |
750 |
751 |
752 |
753 |
754 |
755 |
756 |
757 |
758 |
759 |
760 |
761 |
762 |
763 |
764 |
765 |
766 |
767 |
768 |
769 |
770 |
771 |
772 |
773 |
774 |
775 |
776 |
777 |
778 |
779 |
780 |
781 |
782 |
783 |
784 |
785 |
786 |
787 |
788 |
789 |
790 |
791 |
792 |
793 |
794 |
795 |
796 |
797 |
798 |
799 |
800 |
801 |
802 |
803 |
804 |
805 |
806 |
807 |
808 |
809 |
810 |
811 |
812 |
813 |
814 |
815 |
816 |
817 |
818 |
819 |
820 |
821 |
822 |
823 |
824 |
825 |
826 |
827 |
828 |
829 |
830 |
831 |
832 |
833 |
834 |
835 |
836 |
837 |
838 |
839 |
840 |
841 |
842 |
843 |
844 |
845 |
846 |
847 |
848 |
849 |
850 |
851 |
852 |
853 |
854 |
855 |
856 |
857 |
858 |
859 |
860 |
861 |
862 |
863 |
864 |
865 |
866 |
867 |
868 |
869 |
870 |
871 |
872 |
873 |
874 |
875 |
876 |
877 |
878 |
879 |
880 |
881 |
882 |
883 |
884 |
885 |
886 |
887 |
888 |
889 |
890 |
891 |
892 |
893 |
894 |
895 |
896 |
897 |
898 |
899 |
900 |
901 |
902 |
903 |
904 |
905 |
906 |
907 |
908 |
909 |
910 |
911 |
912 |
913 |
914 |
915 |
916 |
917 |
918 |
919 |
920 |
921 |
922 |
923 |
924 |
925 |
926 |
927 |
928 |
929 |
930 |
931 |
932 |
933 |
934 |
935 |
936 |
937 |
938 |
939 |
940 |
941 |
942 |
943 |
944 |
945 |
946 |
947 |
948 |
949 |
950 |
951 |
952 |
953 |
954 |
955 |
956 |
957 |
958 |
959 |
960 |
961 |
962 |
963 |
964 |
965 |
966 |
967 |
968 |
969 |
970 |
971 |
972 |
973 |
974 |
975 |
976 |
977 |
978 |
979 |
980 |
981 |
982 |
983 |
984 |
985 |
986 |
987 |
988 |
989 |
990 |
991 |
992 |
993 |
994 |
995 |
996 |
997 |
998 |
999 |
1000 |
1001 |
1002 |
1003 |
1004 |
1005 |
1006 |
1007 |
1008 |
1009 |
1010 |
1011 |
1012 |
1013 |
1014 |
1015 |
1016 |
1017 |
1018 |
1019 |
1020 |
1021 |
1022 |
1023 |
1024 |
1025 |
1026 |
1027 |
1028 |
1029 |
1030 |
1031 |
1032 |
1033 |
1034 |
1035 |
1036 |
1037 |
1038 |
1039 |
1040 |
1041 |
1042 |
1043 |
1044 |
1045 |
1046 |
1047 |
1048 |
1049 |
1050 |
1051 |
1052 |
1053 |
1054 |
1055 |
1056 |
1057 |
1058 |
1059 |
1060 |
1061 |
1062 |
1063 |
1064 |
1065 |
1066 |
1067 |
1068 |
1069 |
1070 |
1071 |
1072 |
1073 |
1074 |
1075 |
1076 |
1077 |
1078 |
1079 |
1080 |
1081 |
1082 |
1083 |
1084 |
1085 |
1086 |
1087 |
1088 |
1089 |
1090 |
1091 |
1092 |
1093 |
1094 |
1095 |
1096 |
1097 |
1098 |
1099 |
1100 |
1101 |
1102 |
1103 |
1104 |
1105 |
1106 |
1107 |
1108 |
1109 |
1110 |
1111 |
1112 |
1113 |
1114 |
1115 |
1116 |
1117 |
1118 |
1119 |
1120 |
1121 |
1122 |
1123 |
1124 |
1125 |
1126 |
1127 |
1128 |
1129 |
1130 |
1131 |
1132 |
1133 |
1134 |
1135 |
1136 |
1137 |
1138 |
1139 |
1140 |
1141 |
1142 |
1143 |
1144 |
1145 |
1146 |
1147 |
1148 |
1149 |
1150 |
1151 |
1152 |
1153 |
1154 |
1155 |
1156 |
1157 |
1158 |
1159 |
1160 |
1161 |
1162 |
1163 |
1164 |
1165 |
1166 |
1167 |
1168 |
1169 |
1170 |
1171 |
1172 |
1173 |
1174 |
1175 |
1176 |
1177 |
1178 |
1179 |
1180 |
1181 |
1182 |
1183 |
1184 |
1185 |
1186 |
1187 |
1188 |
1189 |
1190 |
1191 |
1192 |
1193 |
1194 |
1195 |
1196 |
1197 |
1198 |
1199 |
1200 |
1201 |
1202 |
1203 |
1204 |
1205 |
1206 |
1207 |
1208 |
1209 |
1210 |
1211 |
1212 |
1213 |
1214 |
1215 |
1216 |
1217 |
1218 |
1219 |
1220 |
1221 |
1222 |
1223 |
1224 |
1225 |
1226 |
1227 |
1228 |
1229 |
1230 |
1231 |
1232 |
1233 |
1234 |
1235 |
1236 |
1237 |
1238 |
1239 |
1240 |
1241 |
1242 |
1243 |
1244 |
1245 |
1246 |
1247 |
1248 |
1249 |
1250 |
1251 |
1252 |
1253 |
1254 |
1255 |
1256 |
1257 |
1258 |
1259 |
1260 |
1261 |
1262 |
1263 |
1264 |
1265 |
1266 |
1267 |
1268 |
1269 |
1270 |
1271 |
1272 |
1273 |
1274 |
1275 |
1276 |
1277 |
1278 |
1279 |
1280 |
1281 |
1282 |
1283 |
1284 |
1285 |
1286 |
1287 |
1288 |
1289 |
1290 |
1291 |
1292 |
1293 |
1294 |
1295 |
1296 |
1297 |
1298 |
1299 |
1300 |
1301 |
1302 |
1303 |
1304 |
1305 |
1306 |
1307 |
1308 |
1309 |
1310 |
1311 |
1312 |
1313 |
1314 |
1315 |
1316 |
1317 |
1318 |
1319 |
1320 |
1321 |
1322 |
1323 |
1324 |
1325 |
1326 |
1327 |
1328 |
1329 |
1330 |
1331 |
1332 |
1333 |
1334 |
1335 |
1336 |
1337 |
1338 |
1339 |
1340 |
1341 |
1342 |
1343 |
1344 |
1345 |
1346 |
1347 |
1348 |
1349 |
1350 |
1351 |
1352 |
1353 |
1354 |
1355 |
1356 |
1357 |
1358 |
1359 |
1360 |
1361 |
1362 |
1363 |
1364 |
1365 |
1366 |
1367 |
1368 |
1369 |
1370 |
1371 |
1372 |
1373 |
1374 |
1375 |
1376 |
1377 |
1378 |
1379 |
1380 |
1381 |
1382 |
1383 |
1384 |
1385 |
1386 |
1387 |
1388 |
1389 |
1390 |
1391 |
1392 |
1393 |
1394 |
1395 |
1396 |
1397 |
1398 |
1399 |
1400 |
1401 |
1402 |
1403 |
1404 |
1405 |
1406 |
1407 |
1408 |
1409 |
1410 |
1411 |
1412 |
1413 |
1414 |
1415 |
1416 |
1417 |
1418 |
1419 |
1420 |
1421 |
1422 |
1423 |
1424 |
1425 |
1426 |
1427 |
1428 |
1429 |
1430 |
1431 |
1432 |
1433 |
1434 |
1435 |
1436 |
1437 |
1438 |
1439 |
1440 |
1441 |
1442 |
1443 |
1444 |
1445 |
1446 |
1447 |
1448 |
1449 |
1450 |
1451 |
1452 |
1453 |
1454 |
1455 |
1456 |
1457 |
1458 |
1459 |
1460 |
1461 |
1462 |
1463 |
1464 |
1465 |
1466 |
1467 |
1468 |
1469 |
1470 |
1471 |
1472 |
1473 |
1474 |
1475 |
1476 |
1477 |
1478 |
1479 |
1480 |
1481 |
1482 |
1483 |
1484 |
1485 |
1486 |
1487 |
1488 |
1489 |
1490 |
1491 |
1492 |
1493 |
1494 |
1495 |
1496 |
1497 |
1498 |
1499 |
1500 |
1501 |
1502 |
1503 |
1504 |
1505 |
1506 |
1507 |
1508 |
1509 |
1510 |
1511 |
1512 |
1513 |
1514 |
1515 |
1516 |
1517 |
1518 |
1519 |
1520 |
1521 |
1522 |
1523 |
1524 |
1525 |
1526 |
1527 |
1528 |
1529 |
1530 |
1531 |
1532 |
1533 |
1534 |
1535 |
1536 |
1537 |
1538 |
1539 |
1540 |
1541 |
1542 |
1543 |
1544 |
1545 |
1546 |
1547 |
1548 |
1549 |
1550 |
1551 |
1552 |
1553 |
1554 |
1555 |
1556 |
1557 |
1558 |
1559 |
1560 |
1561 |
1562 |
1563 |
1564 |
1565 |
1566 |
1567 |
1568 |
1569 |
1570 |
1571 |
1572 |
1573 |
1574 |
1575 |
1576 |
1577 |
1578 |
1579 |
1580 |
1581 |
1582 |
1583 |
1584 |
1585 |
1586 |
1587 |
1588 |
1589 |
1590 |
1591 |
1592 |
1593 |
1594 |
1595 |
1596 |
1597 |
1598 |
1599 |
1600 |
1601 |
1602 |
1603 |
1604 |
1605 |
1606 |
1607 |
1608 |
1609 |
1610 |
1611 |
1612 |
1613 |
1614 |
1615 |
1616 |
1617 |
1618 |
1619 |
1620 |
1621 |
1622 |
1623 |
1624 |
1625 |
1626 |
1627 |
1628 |
1629 |
1630 |
1631 |
1632 |
1633 |
1634 |
1635 |
1636 |
1637 |
1638 |
1639 |
1640 |
1641 |
1642 |
1643 |
1644 |
1645 |
1646 |
1647 |
1648 |
1649 |
1650 |
1651 |
1652 |
1653 |
1654 |
1655 |
1656 |
1657 |
1658 |
1659 |
1660 |
1661 |
1662 |
1663 |
1664 |
1665 |
1666 |
1667 |
1668 |
1669 |
1670 |
1671 |
1672 |
1673 |
1674 |
1675 |
1676 |
1677 |
1678 |
1679 |
1680 |
1681 |
1682 |
1683 |
1684 |
1685 |
1686 |
1687 |
1688 |
1689 |
1690 |
1691 |
1692 |
1693 |
1694 |
1695 |
1696 |
1697 |
1698 |
1699 |
1700 |
1701 |
1702 |
1703 |
1704 |
1705 |
1706 |
1707 |
1708 |
1709 |
1710 |
1711 |
1712 |
1713 |
1714 |
1715 |
1716 |
1717 |
1718 |
1719 |
1720 |
1721 |
1722 |
1723 |
1724 |
1725 |
1726 |
1727 |
1728 |
1729 |
1730 |
1731 |
1732 |
1733 |
1734 |
1735 |
1736 |
1737 |
1738 |
1739 |
1740 |
1741 |
1742 |
1743 |
1744 |
1745 |
1746 |
1747 |
1748 |
1749 |
1750 |
1751 |
1752 |
1753 |
1754 |
1755 |
1756 |
1757 |
1758 |
1759 |
1760 |
1761 |
1762 |
1763 |
1764 |
1765 |
1766 |
1767 |
1768 |
1769 |
1770 |
1771 |
1772 |
1773 |
1774 |
1775 |
1776 |
1777 |
1778 |
1779 |
1780 |
1781 |
1782 |
1783 |
1784 |
1785 |
1786 |
1787 |
1788 |
1789 |
1790 |
1791 |
1792 |
1793 |
1794 |
1795 |
1796 |
1797 |
1798 |
1799 |
1800 |
1801 |
1802 |
1803 |
1804 |
1805 |
1806 |
1807 |
1808 |
1809 |
1810 |
1811 |
1812 |
1813 |
1814 |
1815 |
1816 |
1817 |
1818 |
1819 |
1820 |
1821 |
1822 |
1823 |
1824 |
1825 |
1826 |
1827 |
1828 |
1829 |
1830 |
1831 |
1832 |
1833 |
1834 |
1835 |
1836 |
1837 |
1838 |
1839 |
1840 |
1841 |
1842 |
1843 |
1844 |
1845 |
1846 |
1847 |
1848 |
1849 |
1850 |
1851 |
1852 |
1853 |
1854 |
1855 |
1856 |
1857 |
1858 |
1859 |
1860 |
1861 |
1862 |
1863 |
1864 |
1865 |
1866 |
1867 |
1868 |
1869 |
1870 |
1871 |
1872 |
1873 |
1874 |
1875 |
1876 |
1877 |
1878 |
1879 |
1880 |
1881 |
1882 |
1883 |
1884 |
1885 |
1886 |
1887 |
1888 |
1889 |
1890 |
1891 |
1892 |
1893 |
1894 |
1895 |
1896 |
1897 |
1898 |
1899 |
1900 |
1901 |
1902 |
1903 |
1904 |
1905 |
1906 |
1907 |
1908 |
1909 |
1910 |
1911 |
1912 |
1913 |
1914 |
1915 |
1916 |
1917 |
1918 |
1919 |
1920 |
1921 |
1922 |
1923 |
1924 |
1925 |
1926 |
1927 |
1928 |
1929 |
1930 |
1931 |
1932 |
1933 |
1934 |
1935 |
1936 |
1937 |
1938 |
1939 |
1940 |
1941 |
1942 |
1943 |
1944 |
1945 |
1946 |
1947 |
1948 |
1949 |
1950 |
1951 |
1952 |
1953 |
1954 |
1955 |
1956 |
1957 |
1958 |
1959 |
1960 |
1961 |
1962 |
1963 |
1964 |
1965 |
1966 |
1967 |
1968 |
1969 |
1970 |
1971 |
1972 |
1973 |
1974 |
1975 |
1976 |
1977 |
1978 |
1979 |
1980 |
1981 |
1982 |
1983 |
1984 |
1985 |
1986 |
1987 |
1988 |
1989 |
1990 |
1991 |
1992 |
1993 |
1994 |
1995 |
1996 |
1997 |
1998 |
1999 |
2000 |
2001 |
2002 |
2003 |
2004 |
2005 |
2006 |
2007 |
2008 |
2009 |
2010 |
2011 |
2012 |
2013 |
2014 |
2015 |
2016 |
2017 |
2018 |
2019 |
2020 |
2021 |
2022 |
2023 |
2024 |
2025 |
2026 |
2027 |
2028 |
2029 |
2030 |
2031 |
2032 |
2033 |
2034 |
2035 |
2036 |
2037 |
2038 |
2039 |
2040 |
2041 |
2042 |
2043 |
2044 |
2045 |
2046 |
2047 |
2048 |
2049 |
2050 |
2051 |
2052 |
2053 |
2054 |
2055 |
2056 |
2057 |
2058 |
2059 |
2060 |
2061 |
2062 |
2063 |
2064 |
2065 |
2066 |
2067 |
2068 |
2069 |
2070 |
2071 |
2072 |
2073 |
2074 |
2075 |
2076 |
2077 |
2078 |
2079 |
2080 |
2081 |
2082 |
2083 |
2084 |
2085 |
2086 |
2087 |
2088 |
2089 |
2090 |
2091 |
2092 |
2093 |
2094 |
2095 |
2096 |
2097 |
2098 |
2099 |
2100 |
2101 |
2102 |
2103 |
2104 |
2105 |
2106 |
2107 |
2108 |
2109 |
2110 |
2111 |
2112 |
2113 |
2114 |
2115 |
2116 |
2117 |
2118 |
2119 |
2120 |
2121 |
2122 |
2123 |
2124 |
2125 |
2126 |
2127 |
2128 |
2129 |
2130 |
2131 |
2132 |
2133 |
2134 |
2135 |
2136 |
2137 |
2138 |
2139 |
2140 |
2141 |
2142 |
2143 |
2144 |
2145 |
2146 |
2147 |
2148 |
2149 |
2150 |
2151 |
2152 |
2153 |
2154 |
2155 |
2156 |
2157 |
2158 |
2159 |
2160 |
2161 |
2162 |
2163 |
2164 |
2165 |
2166 |
2167 |
2168 |
2169 |
2170 |
2171 |
2172 |
2173 |
2174 |
2175 |
2176 |
2177 |
2178 |
2179 |
2180 |
2181 |
2182 |
2183 |
2184 |
2185 |
2186 |
2187 |
2188 |
2189 |
2190 |
2191 |
2192 |
2193 |
2194 |
2195 |
2196 |
2197 |
2198 |

```

RESULT 7
 BG012371 299 bp mRNA EST 24-JAN-2001
 LOCUS RC2-GN0280-131200-012-e02 GN0280 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG012371
 ACCESSION BG012371
 VERSION BG012371.1 GI:12461495
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
 Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,
 Brunstein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare
 M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
 Simpson, A. J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=RC2&t2=RC2-GN0280-131200-012-e02&t3=2000-12-13&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 38
 High quality sequence stop: 299.

FEATURES
 source
 1..299
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GN0280"
 /dev_stage="Adult"
 /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
 ; Site_2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. Letters Patent
 application No. 196,716 - Ludwig Institute for Cancer
 Research) profiles into the puc 18 vector. Reverse
 transcription of tissue mRNA and cDNA amplification were
 performed under low stringency conditions."
 59 a 98 c 62 g 80 t
 BASE COUNT
 ORIGIN

Query Match 2.4%; Score 18; DB 172; Length 299;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 367 gaagtgccagggtttcc 384
 |||||
 DB 54 GAAGTGGCCAGGTTCC 71

RESULT 8
 A2426939/c 332 bp DNA GSS 03-OCT-2000
 LOCUS 1M0208H08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC1M0208H08 R, DNA sequence.
 ACCESSION A2426939
 VERSION A2426939
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112 USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0208 row: H column: 08
 Seq primer: CACACAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 332.
 Location/Qualifiers
 1..332
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0208H08"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (gil4732114[gb]AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 82 a 88 c 71 g 91 t
 ORIGIN
 Query Match 2.4%; Score 18; DB 243; Length 332;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 154 aggcctggtagaagt 171
 |||||
 DB 49 AGGCCCTGGTAGAAGTG 32

RESULT 9
 AU055708 342 bp mRNA EST 29-APR-1999
 LOCUS AU055708 Oryza sativa mature leaf Nipponbare Oryza sativa CDNA
 DEFINITION clone S20009_1A, mRNA sequence.
 ACCESSION AU055708
 VERSION AU055708.1 GI:4714591
 KEYWORDS EST.
 SOURCE Oryza sativa.

```

ORGANISM      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1 (bases 1 to 342)
AUTHORS        Yamamoto,K. and Sasaki,T.
TITLE          Rice cDNA from mature leaf
JOURNAL        Unpublished (1999)
COMMENT        Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program
                2-1-2 Kannondai,Tsukuba
                Ibaraki
                Japan 305
                Tel: 0298-38-7441
                Fax: 0298-38-7468
                Email: tsasaki@abr.affrc.go.jp
PROJECT        'RGP'
FEATURES       Location/Qualifiers
                1..342
                /organism="Oryza sativa"
                /strain="Nipponbare"
                /db_xref="taxon:4530"
                /clone="520009.1A"
                /clone_lib="Oryza sativa mature leaf Nipponbare"
                /tissue_type="mature leaf"
BASE COUNT    111 a 84 c 66 g 81 t
ORIGIN
Query Match      2.4%; Score 18; DB 107; Length 342;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 agtccaacatcattaccg 226
      ||||||||||||||||
Db 146 AGTCCACATCAATACCG 163

RESULT 10
AI069694/c 362 bp mRNA EST 21-SEP-2000
LOCUS TENU2738 T. cruzi epimastigote normalized cDNA Library Trypanosoma
DEFINITION cruzi cDNA clone 26p14 5', mRNA sequence.
ACCESSION AI069694
VERSION AI069694.1 GI:3392669
KEYWORDS EST.
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE      1 (bases 1 to 362)
AUTHORS        Porcel,B.M., Tran,A.-N., Tammi,M., Nyarady,Z., Rydaker,M., Urmenyi
                T.P., Rondinelli,E., Pettersson,U., Andersson,B. and Aslund,L.
TITLE          Gene survey of the pathogenic protozoan Trypanosoma cruzi
JOURNAL        Genome Res. 10 (8), 1103-1107 (2000)
MEDLINE        20414748
COMMENT        Contact: Aslund L
                Department of Medical Genetics
                Uppsala University
                Biomedical Center, Box 589, S-751 23 Uppsala, Sweden
                Tel: 46 18 471 45 85
                Fax: 46 18 52 68 49
                Email: lena.aslund@medgen.uu.se
                Seq primer: T7
                High quality sequence stop: 362.
FEATURES       Location/Qualifiers
                1..362
                /organism="Trypanosoma cruzi"
                /strain="Cl-Brenner"
                /db_xref="taxon:5693"
                /clone="26p14"
                /clone_lib="T. cruzi epimastigote normalized cDNA Library"

```

```

/cell_type="epimastigote"
/notes="cDNA library constructed with oligo dT primed
epimastigote mRNA and cloned in pT7318D phagemid with
modified polylinker (PHARMACIA)"
BASE COUNT    115 a 47 c 115 g 85 t
ORIGIN
Query Match      2.4%; Score 18; DB 15; Length 362;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 cgttagaagaagtccagt 58
      ||||||||||||||||
Db 195 CGTAGAAGAAGTTCAGT 178

RESULT 11
AI544667 371 bp mRNA EST 22-MAR-1999
LOCUS fb77d09.x1 zebrafish WashU MPIMG EST.Danio rerio cDNA 3', mRNA
DEFINITION sequence.
ACCESSION AI544667
VERSION AI544667.1 GI:4462040
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE      1 (bases 1 to 371)
AUTHORS        Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
                S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
                K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
                Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
                Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
                and Wilson,R.
TITLE          WashU Zebrafish EST Project 1998
JOURNAL        Unpublished (1998)
COMMENT        Contact: Stephen L. Johnson
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: zbrfish@watson.wustl.edu
                cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
                Matthew Clark. DNA Sequencing by: Washington University Genome
                Sequencing Center Clone Distribution: Genome Systems, St. Louis,
                Missouri (web address: www.genomesystems.com) (email contact:
                info@genomesystems.com) and Research Genetics, Huntsville, Alabama
                (web address: www.resgen.com) (email contact: info@resgen.com) and
                RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address:
                www.rzpd.de)
                zebrafish identity (p-value greater than 1e-99) found to:
                gi|2289407|gb|AA542472|AA542472 fa07e08.sl zebrafish ICRFzfls Danio
                rerio cDNA
                Seq primer: T7 ET from Amersham
                High quality sequence stop: 347.
FEATURES       Location/Qualifiers
                1..371
                /organism="Danio rerio"
                /db_xref="taxon:7955"
                /clone_lib="zebrafish WashU MPIMG EST"
                /sex="mixed"
                /tissue_type="26 somite embryos, adult livers, shield
                stage embryos"
                /lab_host="XLI-blue MRF"
                /notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                strand cDNA was primed with a Not I - oligo(dT)15 primer
                [5'pgactagttctagatcgagcgccgctttttttttttt3'];
                double-stranded cDNA was ligated to Sal I adaptors (BRL),
                digested with Not I and cloned into the Not I and Sal I
                sites of the pSPORT1 vector (BRL). Library was constructed

```

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 90 a 94 c 65 g 122 t
ORIGIN

Query Match 2.4%; Score 18; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 gtccagttccattgtcttc 123
Db 196 gtccagttccattgtcttc 213

RESULT 12
LOCUS AW390681/c
DEFINITION CM2-ST0194-181099-021-g07 ST0194 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW390681
VERSION AW390681.1 GI:6895340
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project.
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM2&t2=CM2-ST0194-181099-021-g07&t3=1999-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 429.
Location/Qualifiers
1. 429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0194"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 96 a 117 c 107 g 109 t
ORIGIN

Query Match 2.4%; Score 18; DB 115; Length 429;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches
QY 151 ggaagccctggtagaa 168
Db 232 GGAAGCCCTGGTAGAA 215

RESULT 13
LOCUS AQ886579/c
DEFINITION HS_5544_A1_B05_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=9312 Col=9 Row=C, DNA sequence.
ACCESSION AQ886579
VERSION AQ886579.1 GI:6318046
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 436)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 9312 row: C column: 9
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 436.
Location/Qualifiers
1. 436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 97 a 91 c 86 g 160 t
ORIGIN

Query Match 2.4%; Score 18; DB 235; Length 436;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 cagtttcagtcacacat 218
Db 343 CAGTTTCAGTCCAACAT 326

RESULT 14
LOCUS BE606626/c
DEFINITION WHE0901_G12_M232S Wheat 5-15 DAP spike cDNA library Triticum

Query Match 2.4%; Score 18; DB 115; Length 429;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

ACCESSION   BE506626
VERSION     BE506626.1
KEYWORDS    EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum

REFERENCE   1 (bases 1 to 446)
AUTHORS    Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
           P.S., Haia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
           Seaton,C.L. and Tong,J.C.
TITLE      The structure and function of the expressed portion of the wheat
           genomes - 5-15 DAP spike cDNA library
JOURNAL     Unpublished (2000)
COMMENT    Contact: Olin Anderson
           US Department of Agriculture, Agriculture Research Service, Pacific
           West Area, Western Regional Research Center
           800 Buchanan Street, Albany, CA 94710, USA
           Tel: 5105595773
           Fax: 5105595818
           Email: oandersn@pw.usda.gov
           Sequence have been trimmed to remove vector sequence and low
           quality sequence with phred score less than 20.
           Seq primer: Stratagene SK primer.
           Location/Qualifiers
FEATURES    source
             1..446
             /organism="Triticum aestivum"
             /cultivar="Chinese Spring"
             /db_xref="taxon:4565"
             /clone_lib="WHE0901_G12_M23"
             /tissue_type="Spike"
             /dev_stage="Adult plant"
             /lab_host="E. coli SOLR"
             /notes="vector: Lambda Uni-ZAP XR, excised phagemid;
             Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
             greenhouse. Spikes at 5, 10 and 15 DAP were harvested,
             total RNA and poly(A) RNA were prepared, a cDNA library
             was made, and the cDNA clones were in vivo excised to
             give pBluescript phagemids in the T3 Close lab (Choi,
             Close, Fenton) at the University of California,
             Riverside. Plasmid DNA preparations and DNA sequencing
             were performed in the OD Anderson lab (all other authors
             )."
BASE COUNT   143 a   96 c   111 g   96 t
ORIGIN

Query Match      2.4%; Score 18; DB 137; Length 446;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  476 ttttcaggttcacact 493
    |||||
Db   357 TTTTCCAGTTCATCAACT 340

RESULT 15
BF422269/c
LOCUS      BF422269      448 bp      mRNA      EST      28-NOV-2000
DEFINITION FMI_13_D07_b1_A003 Floral-Induced Meristem 1 (FMI) Sorghum
           propinquum cDNA, mRNA sequence.
ACCESSION  BF422269
VERSION    BF422269.1
KEYWORDS   EST.
SOURCE     Sorghum propinquum.
ORGANISM   Sorghum propinquum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
           clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE  1 (bases 1 to 448)

```

```

AUTHORS      Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
              ,L.H.
TITLE        An EST database from Sorghum: floral-induced meristems
JOURNAL      Unpublished (2000)
COMMENT      Contact: Cordonnier-Pratt MM
              Department of Botany
              The University of Georgia
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 542 1805
              Email: mmpratt@uga.edu
              Sequences have been trimmed to exclude polyA, vector and regions
              below Phred quality 16. The threshold for highest quality sequence
              is 20.
              Seq primer: JEN REV
              High quality sequence stop: 379
              POLYA-No.
              Location/Qualifiers
FEATURES     source
              1..448
              /organism="Sorghum propinquum"
              /db_xref="taxon:132711"
              /clone_lib="Floral-Induced Meristem 1 (FMI)"
              /note="Organ: Floral-induced meristems; Vector:
              pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
              EcoRI; mature plants were placed in a growth chamber for
              15 days with 16 hr darkness and 8 hr light (flowering is
              induced by short-day conditions); 16 days after being
              returned to the greenhouse under natural long days during
              late April/early May, meristems were harvested The
              library was made from poly-A RNA in the cloning vector
              lambda ZAP II. Clones to be sequenced were prepared by
              mass excision."
BASE COUNT   142 a   100 c   114 g   92 t
ORIGIN

Query Match      2.4%; Score 18; DB 148; Length 448;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  476 ttttcaggttcacact 493
    |||||
Db   367 TTTTCCAGTTCATCAACT 350

Search completed: August 23, 2001, 14:09:06
Job time: 8638 sec

```

